Issues in the Analysis of Market Power in California

Scott Harvey and William W. Hogan[1]

October 27, 2000

Abstract: Identifying and diagnosing the exercise of market power is often difficult, particularly in tight supply and demand conditions. This task is further complicated in wholesale electricity markets by the need to take account of reliability policies and their impact on the bidding and operating decisions of even competitive firms. These complexities are exacerbated in California with its acknowledged flaws in market design and where the market rules combine to operate more like a pay-as-bid pricing system, which further reinforces the need to distinguish market inefficiency from the exercise of market power.

I. Overview

This paper addresses some of the issues that arise in the analysis of the competitiveness of the California market during the summer of 2000. The issues are important for California, but the implications extend well beyond the boundaries of this particular market. The example of the California market is cited in virtually every restructuring policy discussion, and the California market interacts directly with the rest of the electricity market in the Western system. The events have started a process that will produce many attempts to sort out the complicated issues. [2]

Prices have been high in California, surprisingly high. The reaction to these high prices should reinforce efforts already underway to address problematic features in the California market

---

[1] Scott Harvey is a Director with LECG, LLC, an economic and management consulting company. William Hogan is the Lucius N. Littauer Professor of Public Policy and Administration at the John F. Kennedy School of Government, Harvard University. The authors are or have been consultants on electricity market design and transmission pricing, market power or generation valuation issues for American National Power, British National Grid Company, Calpine Corporation, Commonwealth Edison, Constellation Power Source, Detroit Edison, Duquesne Light Company, Electricity Corporation of New Zealand, General Electric Capital, GPU, Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd., ISO New England, National Independent Energy Producers, New England Power, New York Energy Association, New York ISO, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, PJM Office of the Interconnection, Putnam, Hayes & Bartlett, Reliant Energy, San Diego Gas & Electric, Sempra Energy, Southern Energy, TransÉnergie, Transpower of New Zealand Ltd., Westbrook Power, Williams Energy Group, and Wisconsin Electric Power Company. Jasper Camacho, John Jankowski and Jennifer Vanderhart provided research assistance. The views presented here are not necessarily attributable to any of those mentioned, and any errors are solely the responsibility of the authors.

[2] California ISO (CAISO), Department of Market Analysis, "Report on California Energy Market Issues and Performance: May-June, 2000," August 10, 2000, hereafter DMA. Severin Borenstein, James Bushnell and Frank Wolak, "Diagnosing Market Power in California's Restructured Wholesale Electricity Market," August 2000, hereafter BB&W. Frank A. Wolak, Robert Nordhaus, and Carl Shapiro, "An analysis of the June 2000 Price Spikes in the California ISO's Energy and Ancillary Services Markets," September 6, 2000, hereafter MSC. Northwest Power Planning Council, Study of Western Power Market Prices Summer 2000, October 11, 2000, hereafter NPPC. California Power Exchange Corp., Compliance Unit, Price Movements in California Electricity Markets, September 29, 2000 (hereafter CalPXII).

design.[3] The focus of policy has been the greater emphasis on the possible role of market power in contributing to the troubles. The need for policy and design changes in California seems undisputed, but it is important to get the diagnosis correct as to the nature and relative importance of the different sources of market dysfunction. Beyond the high prices themselves, evidence to demonstrate the exercise of market power must come from further more detailed analysis. Since high prices could in principle be caused by factors other than the conventional exercise of market power, analysis is especially important to the extent that it gets below the surface and identifies the root causes of market turmoil. Although the accumulating reports provide a wealth of information and improved understanding about many issues, the core analysis of market power centers on a few theoretical propositions and methodological tools.

In analyzing the competitiveness of the California market it is often proposed that the exercise of the market power can be identified in the California market by applying the criteria that "offering power at a price significantly above marginal production (or opportunity cost), or failing to generate power that has a production cost below the market price, is an indication of the exercise of market power."[4] There are circumstances in which the first criterion for the identification of market power is appropriate but the presence of these circumstances has not been established in California. Rather, California's "unique" market design includes important elements that would predictably cause even a perfectly competitive firm to submit bids that differ greatly from its marginal production costs. This outcome is therefore, at least in part, a consequence of the market design, particularly its many pay-as-bid elements, and can be avoided only by changing the market design. Moreover, pay-as-bid pricing systems introduce inefficiencies that raise market prices in a manner that can be hard to distinguish from the exercise of market power. Given the realities of the California market design, concluding that supplier bids which differ from marginal production cost reveal the existence of market power may divert attention from the real issues.

As for the failure to generate power that has a production cost below the market price as evidence of market power, direct evidence here would be especially important. A significant pattern of plants found not producing energy or providing reserves when their opportunity costs (rather than just engineering production costs) were below the market price would be a powerful indicator of the exercise of market power. However, the evidence so far has not established that this has occurred. It is noteworthy in considering the analysis of the exercise of market power that none of the studies to date has reported any direct data indicating that a significant number of generating units were held off line providing neither energy nor reserves during the high priced hours during June. Moreover, as discussed below, even this criterion must be applied carefully, recognizing start-up costs, the role of reserves and the impact of uncertainty within the California market design.

[3] John Chandley, Scott Harvey and William W. Hogan, "Congestion Management in California," August 31, 2000. Scott Harvey and William W. Hogan, "Comments on the Congestion Management Proposals of the California ISO," August 31, 2000.

[4] See, for example, BB&W p. 5.

The principal evidence provided of the possible exercise of market power is indirect analysis rather than direct identification of plants not producing. In particular, the MSC analysis comes in the form of competitive market simulations and comparison of simulated competitive prices with the observed prices in the California Power Exchange (PX) market. For example, BB&W develop simulation based estimates of market-clearing prices for a competitive California electricity market and find that their simulated market-clearing prices are on average lower than actual PX prices [5] and much lower during June 2000, which they attribute to the exercise of market power by thermal generators inside California.[6] Finding indirectly what we cannot see directly is a hard problem. While these studies reflect a heroic effort to address an important empirical question, on closer examination, the evidence provided by BB&W and the MSC does not dictate the conclusion they reach. There is a series of methodological features of these studies that could cause the simulation results to understate the competitive level of market-clearing prices. Moreover, these methodological choices would be likely to cause the simulation analysis to understate the competitive market-clearing price to a particularly large degree under the demand and supply conditions prevailing in California electricity markets during the summer of 2000.

Several features of the high California prices seem inconsistent with the market power explanation. For example, prices in August were high not just during peak periods, but also during off peak hours with low loads. In addition, the magnitudes of the difference between observed prices and the BB&W simulations of the competitive prices are very large. Moreover, spot prices appear to have risen to similar levels throughout the WSCC, not merely in California. In each case, if market power were the explanation it would imply that the exercise of market power has been both large and pervasive. If true, direct evidence of extensive withholding and many plants that are "failing to generate power that has a production cost below the market price" should be easy to find, especially for the thermal generation that tends to be the focus of market power concerns. In other words, there should be the proverbial "smoking guns" scattered about the landscape. The absence of such immediate evidence does not guarantee that the smoking guns are not there and just well hidden, but it does raise a question about the possibility that the market power explanation might be incorrect, or at least more complicated than it appears.

Although we cannot rule out the exercise of market power by generators, we note below several kinds of evidence that suggest that factors unrelated to the exercise of market power accounted for the level of prices that prevailed in California electricity markets, and elsewhere in the WSCC, during the summer of 2000. Ideally we would replicate the simulation studies and calculate the impact of the methodological issues we identify. We have not done this due in part to the lack of access to the necessary data.[7] Moreover, we have not demonstrated that there is no market power in the California electricity market nor that none was exercised during the summer of

[5] BB&W p. 33.

[6] MSC p. 17.

[7] We have requested the data in the study to facilitate sensitivity analysis and confirmation of the data, but we understand that some of the underlying data may be confidential.

2000.[8] Rather, we have sought to identify some of the questions that would arise in such a further analysis. Based on these questions, we conclude that the various studies do not yet establish that significant market power either exists or has been exercised in the California market. Pending an opportunity to conduct further numerical analysis, we here outline a view of the complex methodological issues that need to be addressed in drawing conclusions based on simulation studies, such as those of BB&W and the MSC.

Finally, our discussion of the two propositions mentioned above will indicate that one of the critical factors affecting California prices during the Summer of 2000 is the way in which NERC Policy 1 and the corresponding WSCC Minimum Operating Criteria are applied by ISOs, including the California ISO, during capacity shortages. These policy statements can be interpreted as instructing system operators that 10-minute reserves must be maintained at any cost during shortages. This policy could have played a major role in most or all price spike events in electricity markets during the last 4 years, including those in California, MAIN, ECAR and NEPOOL. Moreover, while operating reserves are a necessary element of a reliably operated electricity system, every MW of reserves is not infinitely valuable, and this tension between the absolute mandates of Policy 1 and WSCC Minimum Operating Criteria and the economic reality that reserves are simply not infinitely valuable can lead to price spikes and then to the imposition of price caps. Unlike conditions beyond its control such as natural gas prices, emission limits, and regulatory barriers to plant construction, this is an area in which the Federal Energy Regulatory Commission (FERC) can take actions that will yield more efficient pricing outcomes in California next summer, and likely have similar beneficial effects in Eastern electricity markets.

II. Competitive Bidding in the California Electricity Market

A. Overview

Analyses of competition in California electricity markets often begin by proposing that the exercise of the market power can be identified in the California market by applying the criterion that "offering power at a price significantly above marginal production (or opportunity cost), or failing to generate power that has a production cost below the market price, is an indication of the exercise of market power."[9] These criteria for the identification of market power would be appropriate in some circumstances, but their relevance depends on the market rules. These criteria would be most appropriate in day-ahead electricity markets based on three-part generator bids and inter-temporal operating restrictions, locational marginal pricing, simultaneous optimization of the energy and ancillary service markets, unrestricted demand-side arbitrage with perfect foresight, substantial amounts of price elastic load and reserve targets, and complete

---

[8] Indeed, it is widely recognized that absent mitigation, locational market power would exist in several regions. This potential locational market power has been mitigated to date through RMR contracts. We are not aware of any indications by the CAISO or MSC that such transmission congestion or locational market power played any role in the rise in prices in the California market during the summer of 2000.

[9] BB&W p. 5. Similarly, "the presence of market power can be verified by bid prices significantly over the variable costs of many suppliers in the ISO's markets." DMA p. 5.

coverage of the RTO. These conditions, however, are not found in the California electricity market and some (such as perfect foresight) will not be found in any electricity market.

It is therefore essential in analyzing the competitiveness of California electricity markets to separate the effects of market design and bidding under uncertainty from the exercise of market power. With respect to the first criterion, offering power at a price significantly above marginal production cost, it is necessary to recognize that, given the structure of the California electricity markets, offer prices that exceed a generating unit's incremental production cost, if it operates, do not necessarily provide an indication of the exercise of market power. There are two broad kinds of reasons for discrepancies between the costs and supply offers even of generators entirely lacking market power in the California electricity market. First, in energy and ancillary service markets that clear hour by hour based on one-part bids, competitive suppliers that do not expect to be able to profitably operate at anticipated prices would, to the extent that they submit offer prices at all, submit offer prices that exceed their incremental production costs, if they operate.

Second, California electricity markets are, in effect, segmented such that capacity sold into these markets can clear at a number of independently determined prices, rather than at a single market-clearing day-ahead price. This segmentation of the day-ahead markets provides strong incentives for generators entirely lacking market power to submit bids that reflect the expected market-clearing price, rather than their incremental production costs. In effect, the structure of the California markets causes them to clear more like pay-as-bid markets than as markets based on market-clearing prices. All of these structural characteristics are discussed more fully below.

The second criterion, "failing to generate power that has a production cost below the market price," also must be applied with care within the institutions of the California electricity markets, as there are three broad reasons that suppliers entirely lacking market power might fail to generate power having an incremental production cost that is less than the market price. First, a comparison of production costs with market prices must reflect all of the avoidable costs incurred in selling into the market. In particular, comparisons of incremental production cost with the market price would not reflect start-up and no-load costs that could make operation unprofitable even for a one-plant firm entirely lacking market power. Second, an electric system differs from many other markets in that its reliable operation requires that some capacity be held back from use to generate energy in order to maintain reliability of the overall network in the event of a sudden loss of a generating unit or transmission line (i.e. a contingency). Generating capacity that does not generate power because that capacity is directed by the system operator to provide operating reserves or regulation does not reflect the exercise of market power. Third, pricing systems implicitly or explicitly based on price discrimination, such as pay-as-bid markets, would (absent perfect foresight) be associated with a degree of market inefficiency that would cause some generators lacking market power to fail to generate power in circumstances in which that failure may both raise market prices and reduce generator profits. These issues are discussed below in greater detail and then again in section III as they affect the conclusions of the simulation studies.

Importantly, neither the California ISO nor the MSC have identified any non-quick start units as having been held off-line so as to provide neither energy nor reserves. The Northwest Power

Planning Council has made such an evaluation in the Northwest and concluded that there was no withholding by thermal generators in the Northwest.[10]

B. Inter-temporal Arbitrage

Electricity supply in California is bought and sold through a series of forward and real-time markets. Buyers and sellers can enter into either bilateral contracts (including term contracts) or day-ahead PX transactions both of which are scheduled in the CAISO's day-ahead inter-zonal transmission market.[11] Subsequently, buyers and sellers can enter into either bilateral contracts or intra-day PX transactions, both of which would be scheduled in the CAISO's hour-ahead transmission market. Finally, both generators and loads can be buyers and sellers in the real-time imbalance market.

As discussed at length by Borenstein, Bushnell and Wolak,[12] the expected prices in these temporal markets are related, as loads would not be willing to buy at materially higher prices day-ahead than they expect to prevail in real-time nor would sellers be willing to sell at materially lower prices day-ahead than they expect to prevail in real-time. [13] Thus, sellers lacking market power would withhold output from day-ahead markets, even if the day-ahead price exceeded their production costs, if the seller expected even higher prices to prevail in subsequent markets. This process of arbitraging inter-temporal price differences in electricity is analogous to the similar process that occurs in many other markets.

Multiple temporally related markets of this type do not necessarily require that competitive suppliers depart from cost-based bidding strategies for their physical assets, as arbitrageurs could ensure that prices are in balance between day-ahead and real-time markets, or generators could balance day-ahead and real-time prices with market-based bids on a few marginal units, while bidding most of their units into the market on an incremental production cost basis. There are, however, special features of the California electricity market that are likely to motivate departures from cost-based bidding strategies and may cause a large proportion of the generation supply to submit market-based bids into the day-ahead market or to sell power through bilateral contracts. First, as has received a great deal of attention in California, the California market appears to have characteristics, or demand and supply conditions, that cause the proportion of load that clears in

---

[10] See NPPC pp. 5, 35.

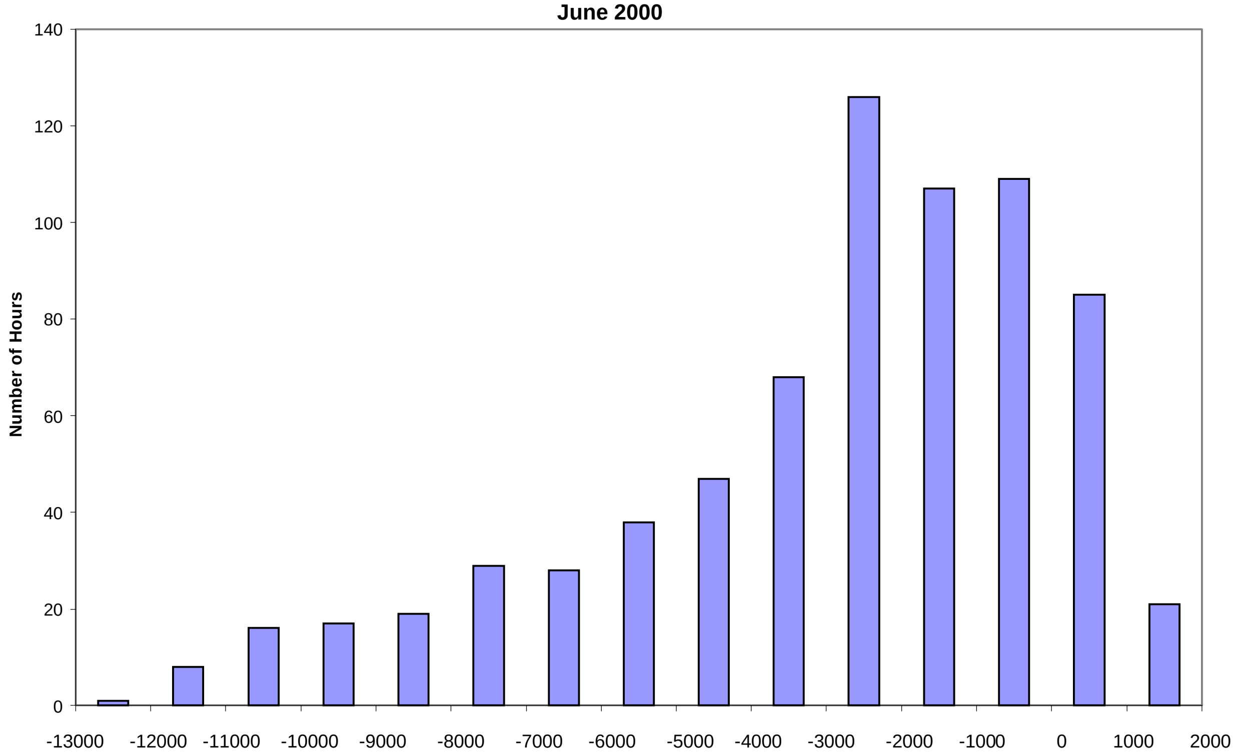
[11] In addition to coordinating market for ancillary services, the CAISO coordinates a day-ahead market for inter-zonal transmission usage. Because there is no day-ahead management of intra-zonal congestion, there is no incentive schedule bilateral transactions within a single zone in the day-ahead market.

[12] BB&W p. 16.

[13] We deliberately use the term "materially" different, because it is possible either that buyers would pay slightly more day-ahead on an expected value basis to avoid the possibility of being exposed to extremely high real-time prices (in the event of a generation outage or incorrect weather forecast) or that sellers would be willing to accept slightly less day-ahead on an expected value basis to avoid the possibility of much lower real-time prices (perhaps attributable to unforeseen changes in weather conditions or imports).

the day-ahead market to be highly variable. Thus, it can be seen in Figure 1 that there were a large proportion of the hours during June 2000 in which the load scheduled day-ahead was within 2000 MWh of real time load, and a large proportion of hours in which real-time load exceeded the load scheduled day-ahead by 4000 MWh or more.

Figure 1
Frequency Distribution: Day-Ahead Load Schedules - Real-Time Load
June 2000



The reasons for this high variability of the proportion of load clearing in the day-ahead market may be related to the binding price caps in real-time, or other considerations. Given this variability, the day-ahead demand curve could intersect a purely cost-based supply curve at a wide variety of prices, which would tend to deter purely cost-based supply offers in the PX. Given that this variability exists, generators lacking market power would quite rationally submit supply offers that reflect the expected market price of energy, to avoid selling energy in the PX for less than the market-clearing price.

A second factor that could push in the opposite direction are the additional costs imposed on loads that buy power in real-time that was not scheduled in the day-ahead market. Third, although generators could arbitrage day-ahead and real-time prices by submitting cost-based supply offers and market-based demand bids, there is little incentive to do so under a one-part bidding system based on zonal pricing. Moreover, PX costs are borne by buyers, so there are material transaction costs to sellers arbitraging day-ahead prices through load bids.

In understanding the level of supply and demand bids, as well as prices, in the day-ahead market, it is important to recognize that weather and other load-affecting conditions are uncertain; real-time prices will therefore sometimes be higher than day-ahead prices and sometimes lower than day-ahead prices. Suppliers attempting to sell at the expected market-clearing price will sometimes submit day-ahead reservation prices that are, in retrospect, too low, that is, the real-time price will turn out to be higher than the day-ahead reservation price they submitted. This was the case for many sellers, for example, on June 13, 14, 21, 26 and 27, as the day-ahead prices were much lower than real-time prices.[14] In these circumstances, the suppliers whose offers were accepted in the day-ahead market would, with the benefit of 20-20 hindsight, wish that they had submitted a higher reservation price in the day-ahead market. At other times, suppliers attempting to sell at the expected market-clearing price will submit day-ahead reservation prices that are in retrospect too high, that is, the real-time price turns out to be lower than the day-ahead reservation price they submitted. This was the case on June 16 and 30, for example. On these days, the suppliers that sold their output in real-time rather than in the day-ahead market would wish in retrospect that they had submitted a lower reservation price in the day-ahead market.

On the other side of the market, buyers make similar decisions, and those buyers who ended up purchasing energy at real-time prices on June 13, 14, 21, 26 and 27 undoubtedly wished in retrospect that they had bought more energy in the day-ahead markets. Conversely, buyers who purchased energy in the day-ahead markets on June 16 and 30 likely in retrospect wished that they had bought less in the day-ahead market and more in real-time.

This arbitrage between day-ahead and real-time markets is a necessary element of electricity markets. The reality that both buyers and sellers will often set reservation prices in the day-ahead markets that turn out to be mistaken does not necessarily reflect the exercise of market power by either loads or suppliers, as long as both loads and suppliers are free to shift their demand and supply between markets.

C. Non-Uniform Pricing

Borenstein, Bushnell and Wolak preface their discussion of the appropriate bidding strategy for a competitive firm in the California market with the observation that: "the California electricity markets, like most electricity markets around the world, clear at a uniform price." [15] This conclusion is not quite correct in two respects, at least one of which is likely to be important in understanding competitive bidding strategy during the summer of 2000. First, it is widely recognized that the California electricity market is based on a pay-as-bid system for intra-zonal congestion management. Thus, generators selected to provide intra-zonal congestion management in real-time are paid their bid rather than the market-clearing price. Pay-as-bid pricing systems will cause perfectly competitive firms to bid their assessment of the market-clearing price at their location, rather than their production costs. While it does not appear that bidding incentives arising from the California intra-zonal congestion management system are likely

[14] See DMA, p. 13.

[15] BB&W p. 5.

to account for the high-priced periods in the summer of 2000, these incentives would likely have influenced bidding over longer periods.[16]

Second, the sequential structure of the energy and individual ancillary service markets operates much like a pay-as-bid pricing system and has effects on bidding incentives that are very much like those of a pay-as-bid pricing system. These bidding incentives are magnified by the "rational buyer" system used for purchasing ancillary services, which also operates like a pay-as-bid pricing system. These incentives are likely to be particularly relevant to understanding bids during the summer of 2000 and may have contributed to the high prices.

An important structural feature of the California electricity market is the split between the PX and the ISO. Generators sell capacity to generate energy in the day-ahead PX market and sell capacity to provide ancillary services in day-ahead markets coordinated by the California ISO. At the margin, there is some capacity that could be used to provide either energy or ancillary services. Suppliers first offer energy portfolios into the PX market (or enter into bilateral contracts), and then after the PX market clears submit unit-specific ancillary service offers to the ISO coordinated ancillary service markets. This sequential structure gives rise to the possibility that rampable capacity on a thermal generator offered into the PX at incremental production cost might be sold at a lower margin [17] in the PX energy market than it would earn had that capacity been sold in the ancillary service markets.

The owners of rampable generation capacity will therefore make an assessment of likely market-clearing prices in the ancillary service markets in offering their capacity into the energy market. Thus, even a perfectly competitive firm with a 5 MW plant capable of ramping from 1 MW to 5 MW in 10 minutes, would not offer its incremental output into the PX energy market at its incremental running cost, but would instead offer this capacity at a price reflecting the higher of its incremental production costs or its expected margin in the ancillary services markets. This offer price would be above the unit's incremental production cost but this difference would not reflect the exercise of market power by a unit with only 5MW of capacity. Instead, such a bid would be entirely consistent with a competitive market, given the structure of the California market. This is recognized by Borenstein, Bushnell and Wolak who observe that ancillary services represent an alternative use, and thus opportunity cost, for capacity that can also be used to generate energy. [18]

This relationship between energy and ancillary service prices has three important implications for analyzing the bidding strategy of competitive firms in the California energy market. First, if there is a shortage in the ancillary service market, generators with capacity capable of supplying either energy or ancillary services will offer that capacity into the energy market at prices reflecting its

---

[16] Intra-zonal congestion has not been mentioned in CAISO discussions of prices during the summer of 2000. See, for example, DMA, p. 5. Intra-zonal congestion did apparently account for the localized load shedding on June 14, 2000, See California's Electricity Options and Challenges, Report to Governor Gray Davis, pp. 21-23.

[17] The term margin as used here means the energy price minus the generation costs that would be avoided by capacity providing reserves.

[18] BB&W p. 16.

expected value in the ancillary services market, which is likely to be above its incremental production cost.[19]

Second, high prices in the ancillary service markets can cause energy-limited generators to establish reservation prices for real-time energy dispatch that are a multiple of the ancillary service prices. Energy-limited generators (i.e. generators with a fixed number of MWh of energy, rather than capacity, available for sale) must recognize that selling a MWh of energy today, might foreclose being able to offer 1 MW of ancillary service for 12 hours the next day. Thus, such a unit with an incremental production cost of $50 and an expected hourly spinning reserve price of $50, might rationally offer its energy into the real-time market at $650.[20] Units of this type could include pondage hydro units with relatively little remaining flexibility to reduce the water level, thermal units that are tightly constrained by emission limits, or gas-fired units that are constrained by a gas shortage.

Third, the sequential energy and ancillary services markets, and pay-as-bid pricing system thereby created, inevitably give rise to bidding mistakes, because competitive firms seeking to earn the market price of their capacity by bidding their capacity into the market at the market-clearing price will at times establish reservation prices in the PX that in retrospect are either too high or too low. The price increases resulting from these bidding mistakes are a result of the pay-as-bid market design, rather than the exercise of market power. The magnitude of these errors cannot be directly measured in the data but some indicators can be examined.

One such measure of the short-term inefficiency of the California market is the number of hours in which the price of energy is less than the price of reserves. In these hours, it is very likely the case that rampable generators offered capacity into the energy market at prices that were too low, given the value of that capacity in the ancillary services market. It can be seen in Table 2 that in June 2000 there were a material number of hours in which such anomalies existed in the day-ahead markets.

---

[19] The pricing of energy and ancillary services in shortage situations is discussed in detail in section II E below.

[20] The forgone ancillary service revenues would be $600 (12*$50) plus $50 incremental production costs.

Table 2
June 2000 Price Relationships

	# of Hours	Percentage of Hours
Energy Price < Spin Price	61	8.47%
Energy Price < 10 Minute Reserves Price	62	8.61%
Energy Price < Replacement Reserves Price	75	10.42%

The energy price is the unconstrained PX price.

Moreover, the data in Table 2 likely understate the degree of disequilibrium in the California energy and ancillary service markets, because there were additional hours in which the price of energy exceeded the price of ancillary services by amounts that were less than the costs avoided by providing ancillary services instead of energy. Table 3 provides a further indicator of the proportion of hours during June 2000 in which bids likely were in disequilibrium due to the structure of the California market by calculating the number of hours in the day-ahead markets in which the energy prices exceeded the various reserve prices by less than $20/MW.

Table 3
June 2000 Price Relationships

	# of Hours	Percentage of Hours
Energy Price < Spin Price + $20	82	11.39%
Energy Price < 10 Minute Reserves Price + $20	80	11.11%
Energy Price < Replacement Reserves Price +$20	82	11.39%

The energy price is the unconstrained PX price.

In addition to the sequential relationship between the energy and ancillary services markets in California, the California ISO clears the various ancillary service markets sequentially, rather than simultaneously, and implements what it calls a "rational buyer" policy in buying ancillary services. First, the sequential process for clearing the ancillary service markets gives rise to the possibility that if suppliers of spinning reserve bid their capacity into the spinning reserve market at cost, they may be paid a lower price for spinning reserve than the market-clearing price of 10 minute or replacement reserves, which their capacity also provides. Thus, generating capacity capable of providing 10 minute spinning reserve could be scheduled by the California ISO to provide spin at a price of $25, in the same hour that the replacement reserve market clears at $250, despite the fact that the capacity providing spinning reserve is also providing replacement reserve. Thus, the sequential clearing process operates like a pay-as-bid pricing system and the effect of these rules is that competitive ancillary service suppliers with capacity capable of providing various kinds of ancillary services would not offer that capacity into the ancillary service markets at incremental cost (or their expected real-time energy margin) but would instead offer this capacity at a price reflecting the expected market price of that capacity. This bidding strategy would have nothing to do with market power, but would be simply an outcome of the pay-as-bid features of the California market rules.

These disequilibrium pricing relationships prevail in the market because sellers have imperfect information and cannot perfectly forecast the market-clearing price of the various ancillary services. It can be seen in Table 4 that there were many many hours in the day-ahead markets during June 2000 in which capacity classified as spinning reserve was paid less than 10 minute reserve and 10 minute reserve was paid less than replacement reserve. It is also noteworthy that the price relationships seen in Tables 2, 3, and 4 do not suggest the exercise of market power but rather that the market is competitive but very inefficient. Entities with market power would not offer spinning reserve at lower prices than they could earn from that capacity in the other reserve markets.

Table 4
June 2000 Reserve Prices

	# of Hours	Percentage of Hours
Spin Price < 10 Minute Reserves Price	205	28.47%
Spin Price < Replacement Reserves Price	93	12.92%
10 Minute Reserves Price < Replacement Reserves Price	173	24.03%

What is not evident in the data is that there were, of course, mistakes in the other direction, in which low cost reserve capacity did not clear in the day-ahead spinning reserve or 10 minute reserve market, because in attempting to bid and earn the market-clearing price, the supplier overestimated the market-clearing price for reserves. These mistakes are a result of the pay-as-bid elements of the pricing system, and lead to artificially high market prices that are a result of inefficiency, not market power.

These disequilibrium and imperfect information effects on bidding incentives are exacerbated by the California ISO's "rational buyer" policy[21] under which the California ISO adjusts the quantities of the various ancillary services that it purchases so as to "substitute lower-cost higher-quality ancillary service bids for lower-quality service to reduce total purchase price." Thus, because ancillary service markets in California clear sequentially and prices are not cascaded, the purchase price of high quality reserves, such as spin, will (as shown in Table 4) sometimes be less than the price of lower quality reserves, such as 10-minute reserves. The rational buyer policy provides that in these circumstances, the California ISO will increase the amount of high quality reserves that it buys at the lower price.

It is important to understand that in these circumstances the market-clearing price of both the high and low quality reserves is the price in the market for the lower quality reserves. The ability of the California ISO to purchase high quality reserves at the lower price arises from the imperfect foresight of suppliers. If suppliers had perfect foresight, no supplier of high quality reserves would offer its capacity into the market for less than the market-clearing price of the lower quality reserves. Moreover, this proposition would hold for generators entirely lacking market power.

The rational buyer policy is therefore a mechanism by which the California ISO attempts to exploit the imperfect information of sellers by purchasing larger quantities of ancillary services at below-market prices. The California ISO claims that the rational buyer model has enabled the ISO to pay much less than the market-clearing price for large amounts of ancillary services "saving" large amounts of money for consumers. Thus, the CAISO recently reported that "savings" under the rational buyer protocol rose to $33.6 million during June 2000.[22] This figure is the amount by which generators supplying reserves were paid less than the market-clearing price for ancillary services as a result of the CAISO purchasing additional high quality reserves at below market-clearing prices when disequilibrium existed.

These claims of savings ignore the reality that pricing systems that attempt to pay suppliers less than the market-clearing price will cause suppliers lacking market power to modify the way in which they bid so as to ensure that they are paid the market-clearing price, by departing further from cost-based bidding strategies, and bidding the market price. Thus, the rational buyer protocol is another element of the market design that causes the California electricity market to operate a pay-as-bid market, with the associated bidding incentives, inefficiency, and inflated prices.

The overall effect of this market structure is that a supplier lacking market power will not simply offer its generating capacity into the day-ahead market at cost and rely on arbitrage to equate the expected day-ahead and real-time price. The California market structure breaks the day-ahead capacity market up into at least 5 markets (energy, AGC up, spinning reserve, 10-minute reserve and replacement reserve). A firm lacking market power that bids its capacity into these markets at cost will likely not be paid the market-clearing price, see Table 2, 3 and 4. A rational response

[21] Implemented August 18, 1999.

[22] Market Analysis Report for June 2000 and July 1-16, 2000 p. 5.

of firms lacking market power to this market structure is to bid their capacity into the market at the expected market-clearing price, not at incremental cost. Another consequence of this market structure is that firms lacking market power, but attempting to bid so as to defeat price discrimination, will at times fail to sell capacity in the market because they misestimate market prices. In these circumstances, a failure to generate, or provide reserves, and the associated price increase does not necessarily reflect the exercise of market power, it may reflect the inefficiency of the California market design.

D. One-Part Bids

Another important feature of the California day-ahead electricity market is that the scheduling of both energy in the PX and ancillary services by the CAISO are based on one-part bids and unit operation is scheduled independently hour by hour. One of the consequences of a one-part bidding system is that potentially extra-marginal generators entirely lacking market power will likely submit one-part bids that exceed their incremental running cost if they were operating, and this difference may at times be quite large. This bidding strategy is rational for firms lacking market power if the generating plant would not necessarily find it profitable to operate at all, as its avoided costs if it is not scheduled to operate are then potentially much greater than its incremental production cost.

In understanding this outcome, it is convenient to begin by considering generators that expect that the market-clearing price in the day-ahead market will be sufficiently high for them to recover their start-up, minimum load and incremental energy costs over the course of the operating day. These generators would find it rational to bid their energy into the market at incremental production cost (aside, of course, from the considerations discussed above relating to inter-temporal arbitrage and the multiple and separate energy and ancillary service markets).

Now consider a generator that does not expect prices to be high enough in enough hours over the day for the generator to recover its cost of starting up. Such a generator would not rationally bid its capacity into the market at marginal cost if its expected return would be negative. The generator could rationally withhold its capacity entirely from the market if it expected operation to be unprofitable, which would be essentially an infinite bid. Such withholding would not reflect market power but merely the expected poor economics of operating. Alternatively, such a generator could recognize that although it expects that the prices over the day in the market would be too low to support profitable operation, its expectations regarding day-ahead prices are based on imperfect information and day-ahead prices may be higher than it anticipates, and that if prices are sufficiently higher than it anticipates in some hours, then its operation might be profitable. In this world of imperfect information, one-part bids and independent hourly markets, a generator lacking market power might well find it profitable to offer its capacity into the market even when it expects operation to be unprofitable, but to offer this capacity at prices well in excess of its incremental production cost.

Consider, for example, a 5MW generator with an incremental running cost of $20/MW, and $100 in start up costs. A generator with 5MW of capacity has an insignificant share of the California market and could not benefit from withholding its capacity from that market as it would then earn no revenues. If the owner of such a unit expected day-ahead prices to exceed $20 in 5 hours and to average $22 in those 5 hours, then its expected profits from being scheduled to operate would be negative.[23] It would therefore not make any sense for the unit owner to bid its capacity into the energy market at its incremental production cost, $20, as if the generator's expectations were correct, it would be scheduled to operate and would lose money if it did so.

The generator owner might, however, consider that if it were actually scheduled in the day-ahead market at a price of $35/MW in any hour, and if the real-time price of energy in the other 4 hours were at least $21.25, then the generator would recover its costs by bidding its capacity into the day-ahead market at $35/MW, and then if it is scheduled to operate in the day-ahead market bidding its unsold capacity into the real-time market at its incremental cost.[24] It should further be noted that a competitive generator submitting one-part bids would likely have the lowest expectation regarding net revenues in other hours other than the peak hour (because the other hours would not include the peak hour) and thus a competitive cost-based bid for such an entity would be higher in the peak hour than in other hours. Suppliers in such a market might further analyze the likelihood that if the price were $35 in one hour in the day-ahead market, then expected real-time prices might also be higher than the supplier expects and find that they could justify bidding into the day-ahead market at $32.[25] Bidding by a firm lacking market power under such separate hourly markets with one-part bids is very complex, but there should be no question that firms completely lacking any market power would, at times, find it profit maximizing to submit energy bids in such a market that differ from and exceed their incremental production cost.

The one-part bidding strategies of generator owners with a portfolio of plants would be somewhat different than those of a single-plant firm, but those bidding strategies would also involve complex departures from the actual incremental production costs of the individual units. Instead, the portfolio bids might be a supply curve for various amounts of capacity using various combinations of units at prices that would recover the start-up, minimum load and incremental production costs of all of the offered units. Moreover, the output of units that were not expected to be profitable to operate would very likely be offered at prices in excess of the incremental production cost of those units, if they operated.

Once again, therefore, it is seen that generators participating in the California market that entirely lack market power would submit bids that differ from their incremental production costs. Similarly, a failure to generate at a time at which the market price exceeded the units incremental production costs would not reflect market power, if the market price were not sufficiently high to also recover the units start-up and no-load costs. Moreover, it is quite possible within the California market structure for a generator lacking market power to find it profitable to submit

[23] 5 hours * 5 MW * $22 - $100 - 5 hours * 5MW * 20 = -$50.

[24] ($35 * 5MW + $21.25 * 5MW * 4 hours) – ($100 + $20 * 5MW * 5 hours) = 0

[25] This relationship might exist, for example, if the high day-ahead price were a result of a generation outage that is not known to the bidder, but would raise both day-ahead and real-time prices.

one–part bids that will at times cause it to not be scheduled to generate energy or provide reserves in the day-ahead market despite day-ahead prices that turn out after the fact to be high enough to permit the generator to recover its start-up and no-load costs. Indeed, this is a likely outcome whenever day-ahead prices are higher than expected, perhaps because of generation outages that were not known to some market participants. Once again, this potential inefficiency is a result of the California market design, and does not necessarily reflect the exercise of market power. While these considerations of start-up and minimum-load cost recovery were unlikely to have affected bidding strategy during the very high priced days during June 2000, they would likely enter into bidding strategy in other periods.

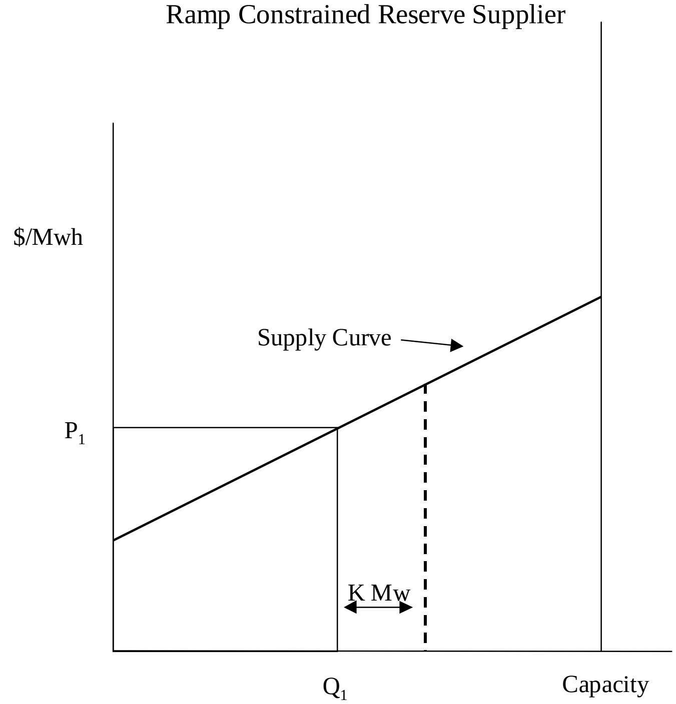
E. Price Determination in Shortages

A final market feature affecting the bidding strategy of competitive firms in the California electricity markets during the summer of 2000 is that the marginal energy and ancillary service supplier is effectively paid its bid rather than the market-clearing price in shortage situations.

1. Market Prices in a Shortage

Before turning to a discussion of bidding strategy in a shortage, it is useful to briefly discuss the relationship between energy and reserve prices in competitive markets. Consider first a situation in which there is more than adequate capacity to meet load and to meet 10 and 30-minute reserve targets. In this situation, there will be generators operating that are both generating energy at a bid price that equals the market price of energy and have undispatched capacity that exceeds their 30-minute ramp capability. This is illustrated in Figure 5, in which P_1 is the market price of energy and K is the quantity of reserves that can be carried on this unit given its ramping capability. The cost of meeting an incremental MW of load with energy from such marginal generators is equal to their energy bid, because increasing their output by 1 MW does not change the amount of reserves they provide (because they are ramp limited, not capacity limited in providing reserves, they can provide the same amount of reserves at the same cost, regardless of their energy output). Thus, from the standpoint of such a generator there is no opportunity cost associated with generating an additional MW of energy, and from the standpoint of the system operator the change in the total production cost of meeting an additional MW of load is equal to that unit's incremental energy bid.

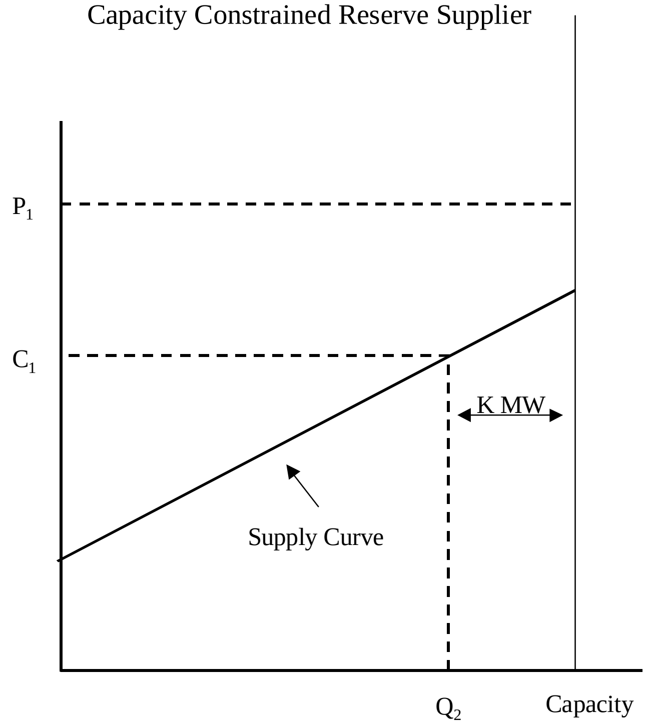
Figure 5
Ramp Constrained Reserve Supplier



At the same time, it could be the case that capacity that is rampable in 10 and 30 minutes could be scarce, and the system operator could back down some in-merit generation such that its energy bid is less than the market price of energy, thereby incurring an opportunity cost of obtaining those marginal reserves. This is illustrated in Figure 6. The capacity of units with such opportunity costs of providing reserves will necessarily be fully utilized in generating energy and providing reserves, such that generating another MW of energy on such a unit requires that the system operator acquire another MW of reserves elsewhere on the system. Thus if the unit in Figure 6 had a 30-minute ramp rate of K MW, All of its capacity would be utilized either to generate energy or to provide reserves. Because the price of energy (P_1) exceeds its generating cost, the generator portrayed in Figure 6 would not provide reserves unless the price of reserves were at least equal to its opportunity cost, $P_1 - C_1$. In competitive equilibrium, the cost of generating another MW of energy on such fully dispatched units that incur opportunity costs in providing reserves will be greater than or equal to the market price of energy, recognizing that the cost of generating energy on such a unit is equal to the incremental production cost of the energy and the cost of acquiring another MW of reserves.[26]

[26] These relationships are described more fully in Michael D. Cadwalader, Scott M. Harvey, William W. Hogan and Susan L. Pope, “Reliability, Scheduling Markets, and Electricity Pricing,” May 1998.

Figure 6
Capacity Constrained Reserve Supplier



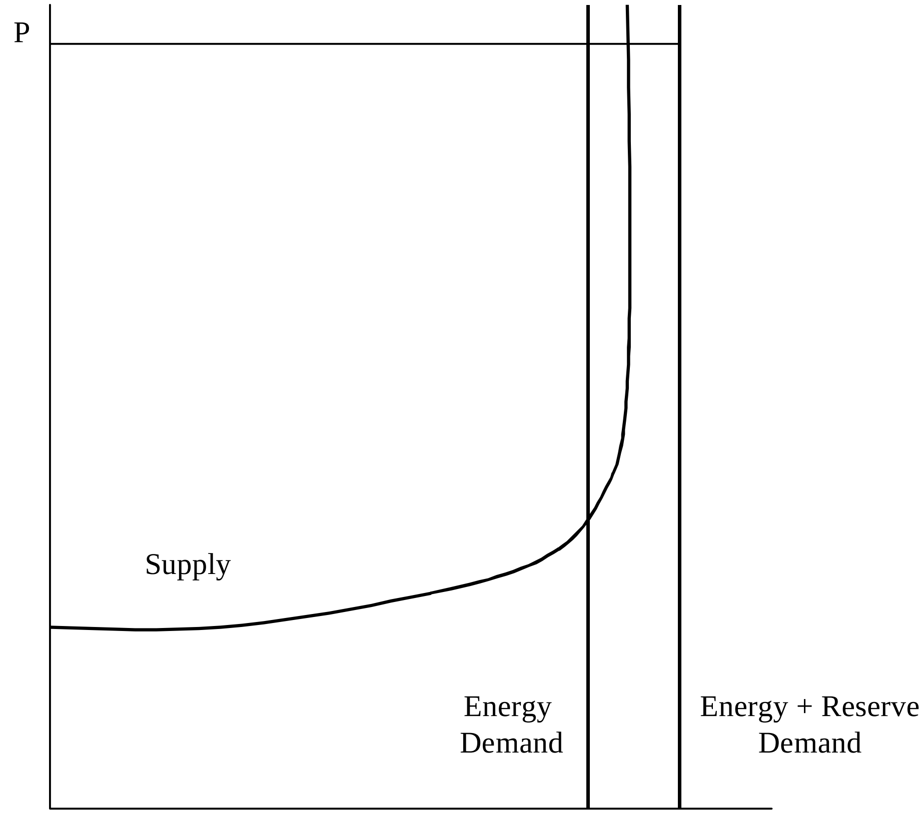
Now consider a circumstance in which there is not enough capacity available in the control area to meet reserve targets without buying reserves from another control area at a price of $750/MWh. In this circumstance, the market price of reserves is $750/MWh. Importantly, many generators will be dispatched like the generator in Figure 6, and the competitive market price of energy in these circumstances will be more than $750/MWh. To see this, let us suppose that the incremental generating cost of the lowest cost unit that is not fully dispatched (C_1 in Figure 6) is $50/MWh. From the standpoint of the system operator, the change in the total production cost of meeting load by dispatching the $50/MWh generator is $800, $50 in incremental energy costs and $750 to buy another MW of emergency reserves (because dispatching the $50/MWh energy reduces reserves by 1 MW). Similarly, from the standpoint of the generator, the cost of providing the energy is $800, $50 in additional fuel and O&M costs and $750 in forgone ancillary service revenues.[27]

The competitive outcome would be the same if the system operator established a $750/MW violation cost for its reserves target. Such a policy would imply that the system operator would be willing to pay up to $750/MWh for reserves and would attach a $750/MWh cost to every MW by which it violated its reserve target. In a shortage situation, the price of reserves would rise to $750/MWh, and the price of energy would be set by the sum of the reserve price and the incremental production cost.

[27] Alternatively, the system operator might restore reserves by purchasing emergency energy from another control area at a price of $800/MWh. By purchasing this energy, the system operator could reduce output by 1MW on internal generators creating 1MW of reserves. The price of energy would be $800/MWh and the price of reserves would be $800/MWh less the avoided costs of the unit whose output is reduced to create reserves.

Figure 7 portrays such a shortage situation from the standpoint of the control area as a whole. It will necessarily be the case that the incremental production cost of the highest cost internal resource generating energy will be less that the price of energy, because there is a large opportunity cost of using capacity to generate energy rather than provide reserves.

Figure 7
Supply and Demand in a Reserve Shortage



It is important to recognize that these relationships between energy and ancillary service prices will arise from the choices of competitive suppliers in the market, regardless of the market design of the ISO-coordinated markets. Thus, even though California's market design attempts to separate energy and ancillary services markets and determines the prices separately, the relationships described above will nevertheless govern the behavior of market participants. In a capacity shortage, ancillary service prices will be high, and these high ancillary service prices will be reflected in energy prices. In shortage hours, the price of energy will reflect both incremental generating costs and the capacity price. Comparisons of the price of energy in shortage situations to incremental generation costs will always therefore find that energy prices appear to be "too high," but this would not reflect the exercise of market power, but only a mistaken criterion for the competitive level of energy prices.[28] While the California market design introduces a lot of

[28]In a shortage situation, one would expect incremental generating costs to be reflected in the difference between the price of energy and the price of reserves.

noise and inefficiency into this relationship between energy and ancillary service prices, the market design cannot undo the underlying economics.

2. Supply and Demand for Capacity in California

In periods in which the California energy and ancillary markets do not clear, i.e. there is not enough capacity available to meet load plus meet all reserve and regulation requirements, the prices in these markets are set by the highest accepted supply offer rather than a shortage price. Since all bids would be accepted in some market in a shortage situation, competitive suppliers have an incentive to submit above-cost bids in an expected shortage situation in order to ensure that they are paid the market-clearing price of capacity. This same incentive exists to a degree during shortages in New York, PJM and NEPOOL. [29] The incentive to bid the market-clearing price is particularly strong in New England [30] and California because the energy market and each ancillary service market clear separately, and these markets operate much like pay–as-bid markets.

In this respect, it is important to understand what is meant by a shortage. Throughout June, the CAISO was able to meet load and the load shedding that occurred was a result of local problems, not a statewide generation shortage. There was therefore no statewide shortage as measured by whether the California ISO was able to meet load. There was nevertheless a generation shortage, as the California ISO was not able to reliably meet load under NERC and WSCC standards. NERC Policy 1 requires that system operators maintain 10-minute reserves equal to the largest single contingency.[31] The WSCC Minimum Operating Criteria impose a similar but even stricter requirement,[32] and this requirement is embodied in various California ISO tariffs,[33] protocols and

---

[29] In New York, however, the day-ahead energy and ancillary service prices are linked, so if there were in fact a shortage, a single high priced bid could set all of the energy and ancillary service prices at high levels.

[30] This comment applies to the market rules that are currently in effect. The New England market participants and ISO-NE intend to eliminate this incentive in both the long-run and short-run market reforms.

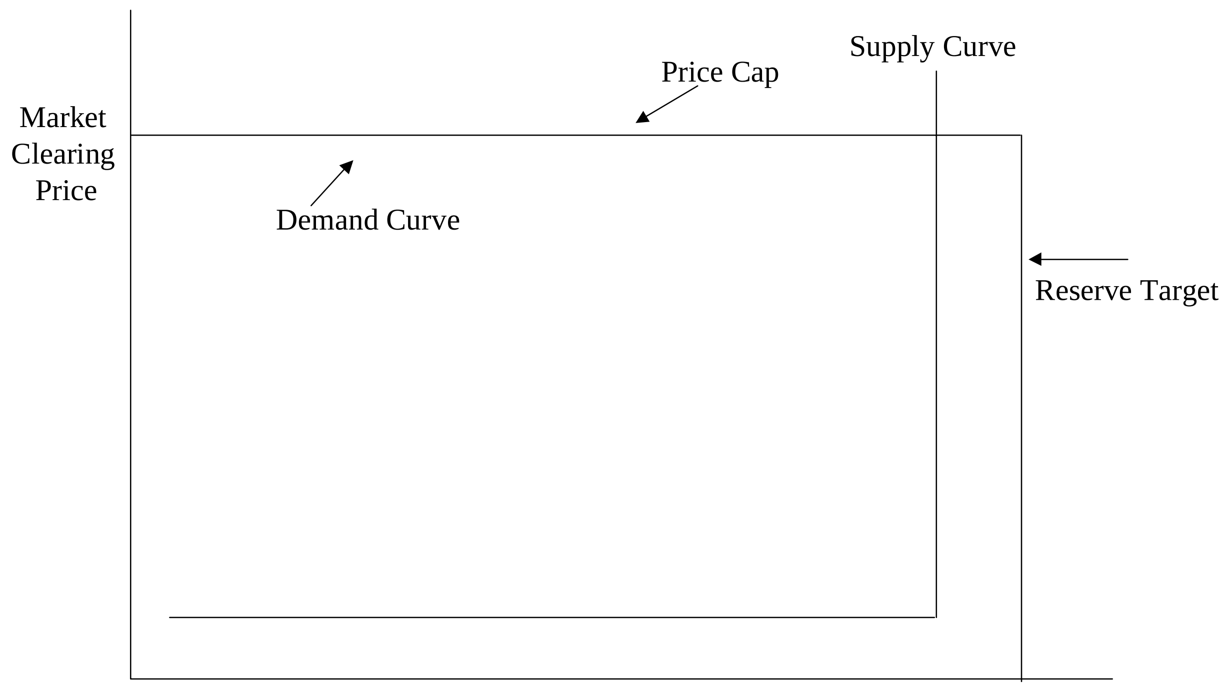
[31] NERC Policy 1 p. P1-1 to P1-2, "Each Region, subregion or reserve sharing group shall specify, and each control area shall provide, as a minimum, operating reserve as follows: (1) Regulating Reserves. An amount of spinning reserve, responsive to AGC, which is sufficient to provide normal regulating margin, plus; (2) Contingency Reserve. An additional amount of operating reserve sufficient to reduce area control error to meet the Disturbance Control Standard following the most sever single contingency. (2.1) Spinning Reserve. At least 50% of this operating reserve shall be spinning reserve, which will automatically respond to frequency deviations."

While California is a large system, its largest contingency is a material amount of generation.

[32] Western Systems Coordinating Council, Minimum Operating Reliability Criteria "Each control area shall maintain minimum operating reserve which is the sum of the following: Regulating Reserve. Sufficient spinning reserve, immediately responsive to automatic generation control (AGC) to provide sufficient regulating margin to allow the control area to meet NERC's Control Performance Criteria. Contingency Reserve. An amount of spinning and nonspinning reserve, sufficient to meet the Disturbance Control Standard as defined in 1.E.2 (a). This Contingency Reserve shall be at least the greater of: (1) The loss of generation capacity due to forced outages of generation or transmission equipment that would result from the most severe single contingency (at least half of which must be spinning reserve); or (2) The sum of five percent of the load responsibility served by hydro generation and seven percent of the load responsibility served by thermal generation (at least half of which must be spinning reserve)." P.2.

operating procedures.[34] What, then, is the competitive market-clearing price in a shortage situation in which the CAISO cannot acquire sufficient reserves to meet the NERC Policy 1 or WSCC standards? In these circumstances, the market-clearing price is the price cap price, because the CAISO is required by NERC and WSCC policy to accept all the supply that is offered at this price. In effect, NERC Policy 1 and WSCC reliability criteria require the ISO to apply a vertical demand curve and the price cap causes the demand curve to become horizontal at the price cap level. The market-clearing price is therefore the price cap price, as shown in Figure 8.

Figure 8



This circumstance is effectively no different than would prevail under any other system of price controls. In a shortage, the market-clearing price will rise to the price cap. Because sellers are not automatically paid the price cap price in a shortage, at least one supplier must bid that price in each product category to set the market price at the price cap level, even in a shortage situation. If all the energy and ancillary service markets cleared jointly, the market would clear at the price-cap price in a shortage situation if only a few suppliers bid the price cap price for a small amount of high cost capacity. Because the various California electricity markets clear separately, there is a potential for suppliers bidding less than the price-cap price in a shortage situation to be selected

[33] The CAISO tariff provides that: "The ISO shall exercise Operational Control over the ISO Controlled Grid to meet planning and Operating Reserve criteria no less stringent than those established by WSCC and NERC as those standards may be modified from time to time, and Local Reliability Criteria that are in existence on the ISO Operations Date and have been submitted to the ISO by each Participating TO pursuant to Section 2.2.1(v) of the TCA." CAISO Tariff section 2.3.1.3.1 sheet 29.

[34] See ISO Operating Procedure, E-504 and E-508.

to sell their capacity in a market in which they are paid less than the market-clearing price of capacity. This market segmentation therefore gives rise to the possibility that some markets may clear below the price cap even in shortage hours because some suppliers underestimate the market-clearing price.[35] While this may superficially appear attractive from the standpoint of loads, this market segmentation can also elevate prices when suppliers make mistakes in guessing the market-clearing price in the other direction. The potential for being selected to provide a product that is paid less than the market-clearing price, i.e. the price-cap price, in a shortage, tends to incent even sellers lacking market power to bid the market-clearing price, rather than their costs, in a likely shortage situation. This incentive is likely to be particularly important with respect to bids in the day-ahead and hour-ahead markets between energy and ancillary services.

Thus, a final circumstance in which generators in the California market might submit bids that are higher than their incremental production cost would be if they anticipated that there would be a shortage of capacity and thus that the ancillary services markets would clear at the price-cap price, thereby also raising to the price cap the expected opportunity cost of using that capacity to generate energy. In order to suggest the empirical relevance of this proposition, we have analyzed the level of available reserves, on a real-time and expected basis, for each hour in June in which the price of energy exceeded $200 and $700. We have attempted to identify the shortage hours using four related approaches. First, we have identified the hours in which there was a shortage of capacity in real-time, signified by the ISO declaring a stage 1 emergency. A stage 1 emergency notice is issued by the ISO in real-time if an operating reserve shortfall is unavoidable or the operating reserve is forecast to be less than the ISO minimum.[36] It is seen in Table 9 that of the 94 hours in June in which the real-time price exceeded $200/MWh in either SP15 or NP15, 38 were stage 1 emergency hours in which the ISO was short of capacity. Similarly, of the 51 hours in June in which the real-time price exceeded $700/MWh in either SP15 or NP15, 31 were stage 1 emergency hours.

Table 9
Real-Time Prices and Stage 1 Emergency Hours
June 2000

	Total Hours	Hours Stage 1 Emergency	% Stage 1
Real-Time Price Exceeds $200/MWh	94	38	40.4%
Real-Time Price Exceeds $700/MWh	51	31	60.8%

[35] For example, the real-time price was only $127.10 for the 10am hour June 27 despite it being declared a first stage emergency.

[36] ISO Operating Procedure E-504.

Table 10 portrays the number of high-priced hours in which the CAISO had issued a warning notice because the capacity available in the hour-ahead market was not sufficient to maintain minimum operating reserves.[37] The issuance of these warning notices permitted the ISO to acquire resources on a non-competitive basis. It is seen in Table 10 that warnings were in effect during 27 of the 28 hours for which warning data are available in which prices exceeded $700/MWh and 38 of the 53 hours for which warning data are available in which prices exceeded $200/MWh. [38]

Table 10
Real-Time Prices and Warning Hours
June 2000

	Total Hours	Warning Hours	% Warning
Real-Time Price Exceeds $200/MWh	53	38	71.7%
Real-Time Price Exceeds $700/MWh	28	27	96.4%

As an additional check on these indicators of shortage conditions, we attempted to compare the capacity available in the hour-ahead market to the real-time demand for capacity by adding a measure of the CAISO's Minimum Operating Reliability Criteria (MORC) minimum operating reserve. It does not appear, however, that the CAISO publicly reports its MORC on an hour by hour basis so a definitive comparison has not been possible. As an approximate check, the 10 minute reserves and upward regulation scheduled in the hour ahead market have been taken as a measure of MORC minimum operating reserve and added to real-time peak load to measure the demand for capacity. Table 11 summarizes the proportion of high-priced hours that were shortage hours based on this comparison. It is evident that more than 85% of the high priced hours in real-time were shortage hours based on this measure of shortage.

[37] ISO Operating Procedure E-504.

[38] Some hours included in Table 9 are omitted in 10. June 13, 14 and 21 are identified in CAISO reports as warning days but no log of warning hours is available for these days. These days are therefore excluded from the totals in Table 10.

Table 11
Real-Time Prices and Shortages
June 2000

	Total Hours	Hours Hour-Ahead Cap < Real-Time demand	% Shortage Hours
RT Price > $200/MWh	93	80	86.0%
RT Price > $700/MWh	51	45	88.2%

Fourth, we have examined whether given the CAISO's day ahead load forecast, the CAISO and other market participants would have expected the market to be in shortage in real-time, by comparing the actual capacity available to the CAISO in the hour-ahead markets to the sum of the CAISO's day-ahead load forecast and hour-ahead 10-minute reserves and upward regulation requirement. Because this shortage measure reflects day-ahead expectations, it is compared to day-ahead PX prices. It is seen in Table 12 that more than 90% of the high priced hours in the day-ahead market were shortage hours by this measure, and all of the hours in which the day-ahead price exceeded $700 were shortage hours.

Table 12
Day-Ahead Prices and Shortages
June 2000

	Total Hours	Hours Day-Ahead Demand > Hour-Ahead Capacity	% Shortage
PX Price > $200/MWh	102	94	92.2%
PX Price > $700/MWh	17	17	100%

The PX price is the unconstrained price.

The California ISO Department of Market Analysis also provided data on the degree of shortage during a number of the high priced days during June. This data also appears to indicate that there was a shortage of capacity during parts of the day on June 13, 14, 15, 26, 27, 28 and 29 even if all of the available non-utility generator capacity had been bid into the market at incremental production cost, and including out-of-market purchases, supplemental import energy bids and replacement imports in the supply and assuming that the ISO sought to acquire total reserves equal to load plus 10%.[39] Moreover, the DMA data appears to indicate that the demand for capacity exceeded the total capacity of the non-utility generators, IOU hydro and must-take energy and scheduled imports during a large number of hours on these days.[40] These data suggest

[39] DMA p. 49.

[40] DMA p. 49.

that energy prices were set by import supply during many of the high-priced hours during June and that the level of demand for capacity depended in considerable part on the level of CAISO purchases of replacement reserves from external sources.

These analyses cannot establish whether all of the high priced hours were in fact shortage hours, because they cannot distinguish hours in which a shortage existed at the hour-ahead price, but the ISO eliminated the shortage through additional non-competitive purchases of import energy or reserves (the $750 imported capacity in the metaphor with which we began this discussion), from hours in which there was no expected shortage. Nevertheless, the data are highly suggestive that the high prices in June arose significantly from a capacity shortage.

The existence of capacity shortages in California during June is consistent with a competitive origin of the high prices but these capacity shortages also do not rule out the exercise of market power. Thus, if non-quick-start capacity were held off line,[41] providing neither energy nor reserves, then that physical withholding could contribute to a shortage and give rise to high prices. None of the data we have analyzed rules out such withholding, but there also does not appear to have been any evidence provided of such capacity withholding.[42] If essentially all of the thermal generation in California was on line [43] and was used to either generate energy or provide reserves, then the source of the high prices was not market power of thermal generators within California but some combination of shortages of capacity and possibly market power exercised by non-thermal generators. If this is the correct perspective, the key ingredient driving high prices in June may have been high demand outside California coupled with a historically poor year for hydro power.[44]

3. Solutions

While the existence of serious capacity shortages in California during June may be sufficient to account for the observed level of energy and ancillary service prices without reference to market power, the discussion above does raise issues that FERC may want to address. The NERC Policy 1 and similar WSCC policies instruct the CAISO to maintain reserves at any cost. The price caps, on the other hand, say, in effect, "No, do not buy reserves at any cost, do not pay more than the price cap because it is not worth more than $250 or $500 or $750/MW to maintain reserves in a shortage." There is a fundamental contradiction here. Rather than indirectly confronting this contradiction by capping prices, it would be better to resolve the contradiction. It is likely that the value of reserves actually depends on the extremity of the shortage, and that the appropriate energy price depends on fuel prices. Thus, a price of $750/MW for reserves might send the

[41] Quick start capacity, such as gas turbines, can provide 10-minute reserves even if not operating.

[42] The California ISO, DMA and MSC have access to data on unit availability and could determine whether capacity not out due to forced outages was held offline. There has been no indication that they found any evidence of such withholding that required further investigation.

[43] Some units will always be unavailable due to forced outages.

[44] See CalPXII p. 21 and NPPC pp. 17-21.

correct price signal to loads and investors in extreme shortages, while it might not be rational to pay more than $25/MW for incremental reserves in very minor reserve shortages, at the same time that energy prices are free to rise far above $25/MWh.

If this view were accepted, rather than confronting the contradiction by imposing price caps that apply in all circumstances, and that apply regardless of the price of fuel and the circumstances in gas, oil, and emissions markets, it would be better public policy for the FERC to coordinate system operator policies regarding the value of reserves in shortage situations. It is widely recognized that a lack of demand elasticity for energy promotes high prices, and the same principle applies to reliability standards in reserve markets. Reserves are not infinitely valuable in all circumstances.[45] NEPOOL and ISO-NE previously proposed the implementation of a demand curve for reserves that would in effect vary the maximum amount that the system operator would pay for a MW of reserves based on the degree of shortage.[46] While the FERC commented favorably on some aspects of the proposal, the FERC order appears to envision that the system operator would continue to meet the NERC requirement in real-time at any price.[47] Deferring the requirement to purchase reserves at any price from day-ahead to real-time would not solve the fundamental problem. Although favorable changes in demand that eliminate forecasted shortages between day-ahead and real-time may obviate the need to buy high-priced reserves, the forecasted shortage may not go away, and reserves not purchased day-ahead at high prices may not be available in real-time, requiring the purchase of other resources at extreme prices. The requirement that system operators maintain reserves in real-time at any price must itself be addressed or system operators will inevitably continue to rely on price caps to avoid irrational outcomes in shortages. Moreover, it would be possible for system operators to implement a flexible target for reserves in shortages without the long delays that may accompany the development of price-sensitive retail load.

F. **Implications for Bidding and Supply**

The considerations discussed in sections C and E above regarding non-uniform pricing and pricing in shortages will likely tend to cause competitive firms to bid above cost in likely shortage situations, with a great many suppliers submitting bids at the price cap. The considerations discussed in sections C and D above regarding non-uniform pricing and one-part bids will likely tend to cause competitive firms to submit bids that are above incremental production cost in non-shortage situations.

---

[45] See, for example, Peter Cramton, Review of the Reserves and Operate Capability Markets: New England's Experience in the First Four Months, Nov 17, 1999

[46] See Appendix C to Fifty-Second Agreement, sections14A5(b) and 14A7(c).

[47] ISO New England Inc, Order Conditionally Accepting Congestion Management and Multi-Settlement Systems, June 28, 2000 p. 14, "The detailed proposal should describe, for example, the exact derivation of the demand curves and how the ISO will meet the NERC requirements in real time in the event that the demand curves result in procuring fewer reserves in the day-ahead market."

Overall, the California market design provides very substantial economic incentives for generators entirely lacking market power to bid their capacity into the market at prices above incremental production cost. It has been suggested that " the presence of market power can be verified by bid prices significantly over the variable costs of many suppliers in the ISO's markets,"[48] but this is not the case given the structure of the California electricity markets.

In addition, the analysis above suggests that any analysis of capacity withholding must account for start-up and no-load costs, as well as incremental energy costs, account for capacity demanded in the form of regulation or reserves, and the impact of short-run inefficiency within the California market design.

III. Empirical Analysis of Market Power in California

A. Overview

The Borenstein, Bushnell, Wolak and MSC papers implicitly recognize the potential disconnect between the costs and bids of competitive suppliers in the California market. As discussed above, a simple direct analysis of bids and costs would not be informative and BB&W do not undertake such an examination of the bid data. Similarly, the BB&W analysis does not provide any data regarding the actual supply or withholding of capacity by generating plants (to provide either energy or reserves) in each hour in comparison with energy prices. Instead, the approach in BB&W to the analysis of market power is indirect and the analysis depends on a chain of reasoning and assumptions. Rather than comparing bids and estimated production costs or identifying capacity that was withheld, they undertake a numerical analysis that could be described as asking the question:

> Even if California market rules require competitive suppliers to bid an estimate of the market-clearing price, rather than their costs, did the pricing outcomes in California energy markets approximate an estimate of what would have been the competitive outcome or did those pricing outcomes reflect the exercise of market power?

In effect, this approach simulates the operation of a competitive market (in general without regard to the fine details of the California market design) and estimates the market-clearing price in such a market. This simulated competitive price is then compared with the observed price and the average differences are attributed largely to the exercise of market power. This is a useful approach to gaining insights into a complex market. However, it is important to analyze the simulation results carefully in assessing whether the simulation results suggest the exercise of market power, or alternatively point to differences between the simulated market and the actual market.

In understanding the strengths and weaknesses of this approach it is helpful to provide a high level overview of the methodology. First, the simulated market price is developed using a demand curve based on actual real-time load rather than some measure of day-ahead expected load.

---

[48] DMA, August 10, 2000 p. 5.

Second, the supply of non-thermal generation is based on the actual real-time output of these suppliers rather than a day-ahead supply curve or schedule. Third, the supply of imported power is apparently a hybrid, based on the real-time import quantity but with a slope based on day-ahead bids. Fourth, the simulation addresses the availability of thermal generation probabilistically by repeating the price calculation for each hour many times to account for the random impact of plant outages. This produces an average simulated price. There is a sense, therefore, in which BB&W are simulating real-time competitive prices, using real-time demand and real-time non-thermal supply.

The price to which BB&W compare this simulated price is not the actual real-time price, however, but the day-ahead PX price. The rational for this comparison is that the day-ahead PX prices will reflect buyers and sellers expectations regarding real-time prices. This will of course not be true on a day-by-day basis, as the expectations of both sellers and buyers will often be wrong, but actual day-ahead and average real-time prices should roughly converge if averaged over a sufficiently long period.

The PX price used for comparison with the simulated price is also not the actual PX price used for settlements, but is instead the hypothetical unconstrained PX price; that is, it is a price calculated without regard to transmission constraints. This is consistent with the treatment of thermal generation in the calculation of the simulated competitive price, as transmission constraints are not taken into account in developing the supply curve for thermal generation. These transmission constraints are, however, implicitly reflected in the actual real-time quantities of non-thermal generation and import supplies, which are also included in the simulation supply curve.

Another subtlety in the modeling approach is that the thermal supply curve used in the simulation is not really a real-time supply curve. Real-time prices are influenced by the chaos of real-time events and unit inflexibilities in many ways not reflected in the simulation. For example, the simulation scheduling of thermal units ignores minimum up and down times. In addition, forced outages of thermal units in effect occur in the simulation prior to unit commitment, so that load can always be met in the simulated dispatch with the lowest-cost units, and gas and gas transportation is always purchased at day-ahead prices. These kinds of real-time constraints are, however, reflected in the supply of import and non-thermal supplies which may reflect the impact of real-time constraints.

Because the structure and assumptions of the simulation can be critical to its conclusions, the simulation and its assumptions would ideally be subjected to independent sensitivity analysis to replicate the results or highlight critical assumptions that are relevant in assessing the policy implications. We have not done this, but such a sensitivity analysis should be undertaken. Every such simulation study requires simplifying assumptions, and these studies are no exception. A close reading of the BB&W and MSC studies suggests that these analyses have sufficiently important simplifying assumptions that they do not yet provide convincing evidence of the exercise of market power in California, particularly in light of data pointing to capacity shortages as playing an important role in the high prices.

There are six general empirical and conceptual limitations of these studies that bear on the conclusions regarding the exercise of market power. First, and likely of particular importance during the summer of 2000, the simulation estimates of competitive price levels are based on the assumption that the California electricity markets could be operated, and thus the energy market would clear with no capacity provided for spinning or non-spinning reserves in excess of load. This assumption is inconsistent with the market structure, inconsistent with the operation of the California grid (as well as the other electricity markets with which we are familiar) and likely has a significant impact on the simulated competitive market price estimates. Second, as BB&W note,[49] the conceptual approach has the property that any inefficiency in the California market arising from the separation of the energy and ancillary services markets and associated pay-as-bid structure or one part bidding that results in higher prices would be treated under the BB&W and MSC studies as arising from the exercise of market power, rather than as costs attributable to limitations in the market design. Third, the conceptual approach in the simulation will necessarily understate the competitive market-clearing price in hours in which thermal generators within California are fully dispatched to provide energy and reserves and non-thermal units were on the margin and setting prices. Fourth, the simulation analysis implicitly assumes that there was no transmission congestion affecting the dispatch of non-thermal generators inside California. In practice, there has been both inter-zonal and intra-zonal congestion that has at times prevented the California ISO from dispatching the lowest as-bid cost generation to meet load. Given the methodology of the BB&W and MSC studies, this congestion can cause them to estimate the simulated competitive market price with a different non-thermal supply than that which determined the unconstrained PX price in the real-world. Conversely, the treatment of RMR generation may at other times cause the BB&W and MSC studies to estimate a simulated competitive market price that is too high. Fifth, the simulation analysis assumes that there are no start-up costs, which will necessarily lead to an underestimate of the competitive market-clearing price, as lower costs will reduce prices. Sixth and finally, there are a variety of assumptions in the measurement of costs that are likely to generally cause the simulation to understate the market-clearing price.

The reality is that the analysis is a complex piece of analysis based on sometimes subtle assumptions regarding market interactions. This complexity gives rise to the difficulties outlined above. The combined effect of most, but not all, of these conceptual and empirical issues is likely to push the studies towards finding that actual prices exceed the simulated competitive prices and this difference could be particularly large during high load periods such as June 2000. The magnitude of the bias could be substantial. There is no reason yet to assume it would be insignificant or that most of the difference between the simulated and actual prices is driven by the exercise of market power. The BB&W analysis advances the conversation, but raises further important questions. Based on this evidence alone, there may have been a significant exercise of market power, or there may have been flawed competitive market design interacting with a general shortage of capacity.

[49] BB&W p. 7.

B. Demand and Supply of Capacity

It is likely that the most important methodological element of the BB&W study is the assumption that the California ISO should and could have cleared the energy and ancillary services markets and operated the electric system in real-time without any capacity set aside to provide reserves in the event of a contingency. This is a strong assumption that is at odds with other information. In view of the importance of this assumption, it is useful to quote their reasoning in full:

> Since the ISO is effectively purchasing considerable extra capacity for the provision of reserves, it might seem appropriate to consider these reserve quantities as part of the market clearing demand level. However, with the exception of regulation reserve, as described below, all other reserves are normally available to meet real-time energy needs if scheduled generation is not sufficient to supply market demand. Thus, the real-time energy price is still set by the interaction of real-time energy demand - including quantities supplied by reserve capacity -- and all of the generators that provide real-time supply. Therefore, we consider the real-time energy demand in each hour to be the quantity that must be supplied and capacity selected for reserve services to be part of the capacity that can meet that demand and, as such, to be part of our aggregate marginal cost curve.[50]

The implication of this view is that system operators need not maintain any operating reserves in excess of load at the system peak and thus that the total demand for capacity on the part of load and the system operator is limited to the peak load plus upward regulation margin. Under this assumption regarding the total demand for capacity, BB&W then simulate the competitive market-clearing price of electricity in California. Taken literally, and if this what was actually done in the analysis, these assumptions regarding reserves would materially understate capacity requirements in the California market. The purpose of operating reserves is not simply to meet load uncertainty and allow the ISO to meet load. Instead, the purpose of operating reserves is also to ensure that the electricity system can operate reliably following contingency events.[51] Thus, the system operator seeks to maintain sufficient reserves to not only meet load and maintain the regulation margin but to at least have sufficient reserves to restore the system following the loss of the single largest transmission line or generating unit. Indeed, as discussed above, ISOs in particular feel compelled to meet the strictures of NERC Policy 1 as well as the WSCC Minimum Operating Reliability Criteria (MORC), which many interpret to require that the system operator acquire sufficient reserves to meet load and meet the single largest contingency <u>at any cost</u>.

If the BB&W characterization of reserve practices were correct, the demand for capacity would have been much lower in both California and the rest of the WSCC this summer and prices undoubtedly would have been much lower. Indeed, if their view of reserves were correct the California ISO was negligent for paying extremely high prices for reserves that were not used to meet energy demand and were therefore not needed under the BB&W characterization of electric

[50] BB&W p. 19.

[51] See WSCC p. 2.

system reliability. Although we are not familiar in detail with the operating policies of the California ISO, it appears based on the CAISO tariff, dispatch protocols and operating procedures discussed above, that the CAISO, like the NY, NEPOOL and PJM ISOs, attempts to meet an obligation to maintain 10 minute reserves, spinning and non-spinning, in excess of its peak load. If the California ISO is complying with the WSCC policy quoted above, then it must maintain capacity in reserve equal to somewhere between 5 and 7% of load, in addition to a regulation margin. If so, the simulations of the competitive market in the BB&W and MSC analyses understate the total demand for capacity by 5-7%. This difference could lead to a material underestimate of the competitive market-clearing price, especially during peak hours.

Put another way, this is a critically important assumption because the CAISO should be operating to ensure that it will have sufficient capacity on line to meet load, maintain AGC (regulation) and maintain a few thousand MW of 10-minute reserves. Moreover, if the CAISO interprets NERC Policy 1 and the WSCC Minimum Operating Reliability Criteria as requiring the CAISO to pay any price up to the price cap in order to maintain reserves, then the market-clearing price of capacity in reserve shortage situations will be the price-cap price. Although we have not had access to the details of the BB&W study, it appears likely that this assumption could materially affect the findings of the study for June 2000 given the recurrent capacity shortages that appear to have actually existed periodically during the month, as discussed in section II E above. The MSC presumably has access to data on actual CAISO reserve levels, so it should be straightforward to simulate prices based on the actual demand for capacity, load and MORC: (AGC and reserves).

It is clear from the discussion in section IIE above that the high prices in California this summer were driven in part by capacity shortages and thus a central critical issue in understanding the reason for these high prices is understanding the cause of these capacity shortages. It will be difficult to use the BB&W methodology to satisfactorily address this issue. Even if all of the other complications discussed below were satisfactorily resolved, the BB&W methodology ultimately simulates forced outages based on some very general outage data. A simulation of outage rates might be useful in assessing whether the rates at which California thermal generators were not available to provide energy or reserves were consistent with conventional outage assumptions, but simulations of this sort cannot establish whether there was withholding, simply bad luck in the timing of forced outages, or hard usage that raised outage rates. The BB&W methodology is likely better suited to examining the exercise of market power during non-shortage conditions.

C. Market Inefficiency

We discussed above, in section II B and II C, the various features of the California electricity markets that will cause suppliers entirely lacking market power to bid their resources into the market at the expected market-clearing price rather than at their costs. If these guesses do not change the resources used to meet load, errors in predicting the market-clearing price could raise the prices paid by loads (if generators guess that the market-clearing price will be higher that it actually turns out to be) or could reduce the prices paid by loads (if generators guess that the market-clearing price will be lower than it actually turns out to be). If there were no change in the actual resource cost of meeting load, the errors might cancel out and have no impact on market efficiency or market prices. In practice, however, it is very unlikely in markets that clear based on generators guessing the market-clearing price that load will be met at least cost. Instead, it is likely that some of the generators that over-estimate the market-clearing price would be lower cost than some of the generators whose bids clear in the market. In this circumstance, the bidding errors attributable to non-uniform pricing would not cancel out, because the supply curve would in effect be shifted in as a result of some supply resources not clearing in the market despite being infra-marginal on a cost basis.

The design of the California market is premised on there being other advantages that compensate for this short-term inefficiency. Even if that is the case, however, the short-term inefficiency must be taken into account in empirical analyses of market performance. While reasonable people may disagree over the magnitude of the short-run inefficiency arising from the non-uniform, pay-as-bid pricing elements of the California market design, it is difficult to identify market power by simulating competitive prices under the assumption that this inefficiency does not exist.

It is clear from the data in Tables 2, 3 and 4 above that many suppliers at times underestimated the observed market-clearing price. Symmetry suggests that many suppliers also at times will overestimate the market-clearing price, so there is a potential for the pay-as-bid market structure to give rise to significant inefficiency. The BB&W analysis of supply in effect calculates the market-clearing prices that would arise under a centralized market-clearing process for energy and ancillary services, rather than simulating likely competitive outcomes under the actual California market structure.

BB&W acknowledge that their approach treats all inefficiencies in the market as market power but argue that "for the great majority of these" flaws, "the flaw would be fairly benign if firms acted as pure price takers, rather than exploiting these design flaws to affect the market price." [52] The basis for this view is not indicated, however, and is at odds with the discussion in II C and II D above. Indeed, it appears to us that the reverse might be more likely, that the inefficiency induced by these market flaws would be less important in a market in which substantial market power was exercised and would be greatest in a market with many firms each trying to bid so as to ensure that it is paid the market-clearing price.

BB&W also argue that because their simulations of competitive prices do not differ greatly from the actual market-clearing prices in some months, their market power estimates are not upward biased due to inefficiency. [53] The apparent premise for this conclusion is an implicit assumption that if their simulation of the competitive price is close to or above the actual PX price, then the competitive market price in an efficient market could not be higher than the simulated price. This is not necessarily the case, however, for a variety of reasons. First, it is important to recognize that the actual market price used in the BB&W comparisons is the day-ahead unconstrained PX price, which should reflect expected load conditions in the day-ahead market. As noted above, the simulated competitive price is calculated based on real-time load. The rationale for such an

[52] BB&W p. 7.

[53] BB&W p. 7.

approach is that day-ahead load expectations should reflect the probability distribution of real-time loads and thus prices based on expected loads will over time average out at roughly the level of prices determined by real-time loads. This convergence exists only over time, however, and not on a day-by-day basis. [54]

Because the BB&W estimate of the competitive market price is based on actual real-time loads, while the actual PX price is driven by the expected price in day-ahead markets, variations between day-ahead expectations and real-time realizations attributable to vagaries of weather within any month can cause the simulated competitive price for a given month to be higher than the actual PX price. Thus, in any hour in which the actual load is higher than the load expected in the day-ahead market, the BB&W simulated competitive price would on average be higher than the PX day-ahead market price, simply because the BB&W demand curve would be expected to intersect the supply curve at a higher point than did the real-world day-ahead demand curve, and conversely for days with lower than expected loads. These random differences between actual and simulated prices would presumably average out in tests based on the relationship between the simulated and actual prices over a sufficiently long-period, but the differences will exist in day-by-day and month by month price comparisons. [55] A finding that the actual and simulated prices are similar over several months therefore does not establish that the simulated price is not downward biased, because the bias may be offset by these kinds of random differences in particular months.

Second, although their measure of petroleum fuel costs for peaking units is generally below market, it is above market for most of the months in which the simulated competitive price is cited as comparable to the actual PX price.

Third, even if the BB&W methodology for simulating the impact of forced outages of generating units were centered around the real-world probability distribution, there has been no demonstration that the realizations would converge over the period of a month. Thus, differences in forced outage realizations between the real-world and the simulated world may tend to produce month to month differences in average monthly prices between the real and simulated markets.

Fourth, as discussed further in section III E below and noted by BB&W, the simulated price will tend to overstate the competitive price in months in which large RMR calls were expected (pre Amendment 26) or occurred (post Amendment 26).

Because of the variety of elements in the simulation structure that would produce variations between real-world and simulated prices, it is risky to assume the existence or non-existence of bias due to market inefficiency in the manner suggested by BB&W. As an alternative to such an

---

[54] As discussed in BB&W, the BB&W analysis assumes that day-ahead prices reflect expected real-time loads, rather than only the amount of load bid into the day-ahead market. One of the complexities in the BB&W analysis is that the measure of load is actual real-time load, while the price is not only the day-ahead price, but the hypothetical unconstrained day-ahead price.

[55] It is noteworthy that even in the real-world, day-ahead and real-time prices do not always converge over a month, see BB&W p. 13.

approach, sensitivity analysis could be undertaken, estimating market clearing prices under the assumption that 2%, 5%, 8% and 10% of the infra-marginal thermal capacity became extra-marginal due to inefficiency arising from the structure of the California market. Policy makers could then make their own assessments of whether the price differences likely arose from market inefficiency or other causes.

D. Non-thermal supply

The Borenstein, Bushnell & Wolak and MSC studies take the output of hydro and geothermal resources as given in defining the competitive supply curve and estimating the competitive market price. As they describe, they in effect place the output of the hydro and geothermal resources at the bottom of the supply curve.[56] While, as BB&W correctly observe, this treatment of hydro and geothermal resources is conservative with respect to the exercise of market power by hydro and geothermal resources,[57] it is not conservative with respect to identifying the exercise of market power by thermal unit owners. In particular, this approach will understate the competitive price level in any hour in which hydro or geothermal units were on the margin.

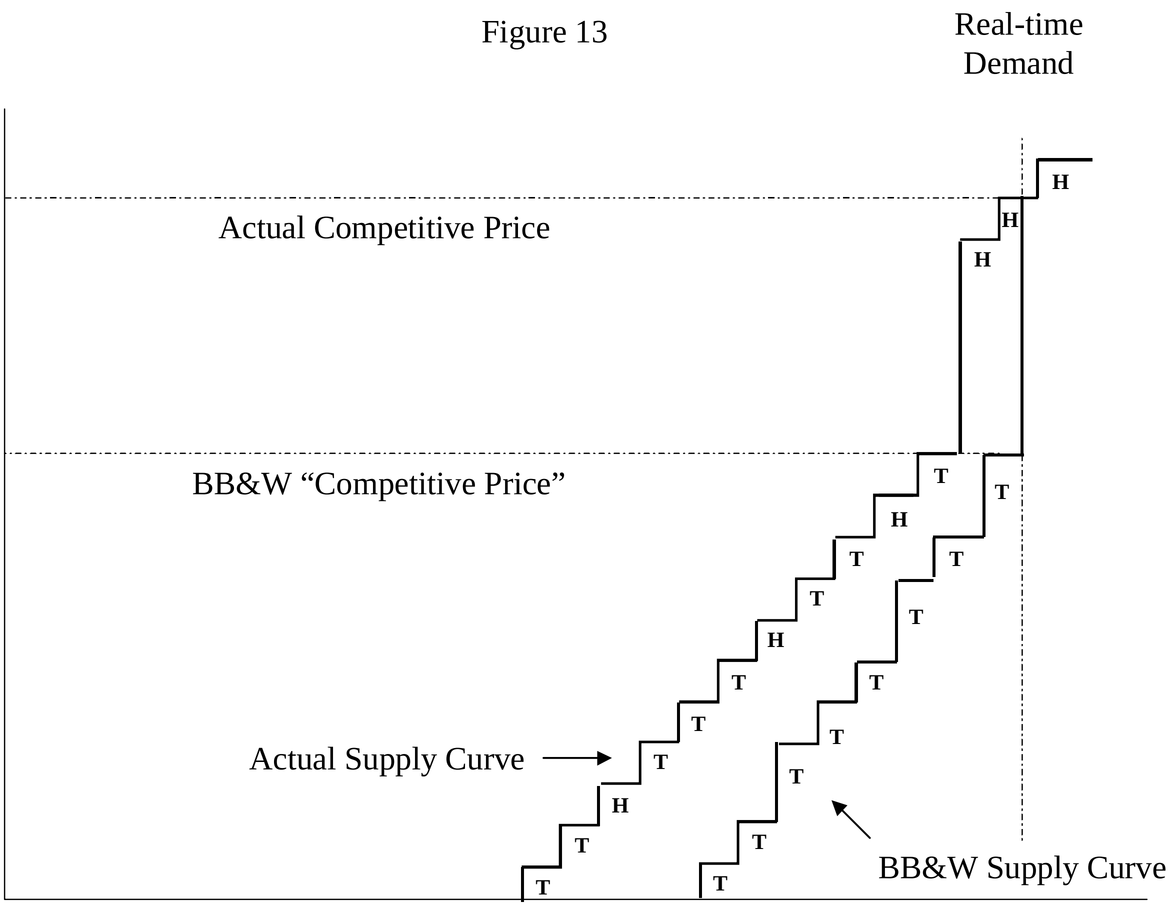
This potential is illustrated in Figure 13. In this figure, high opportunity cost hydro units (labeled H) are on the margin in the real-world, while the thermal units (labeled T) are infra-marginal. The methodology described by BB&W would in effect construct the simulation supply curve by moving the high cost hydro units that were on the margin to the bottom of the BB&W supply curve resulting in the estimated competitive price labeled as the BB&W competitive price, while the PX price determined by exactly the same thermal unit bids would be the much higher line labeled as the "actual" competitive price.

---

[56] BB&W p. 24 and Figure 2.

[57] See BB&W p. 24.

Figure 13



Information regarding the frequency of the time that such hydro and geothermal units were on the margin during June 2000 is apparently not publicly available, but data calculated by the California PX showing that many of the high-price bids in the day-ahead market were submitted for IOU resources suggests that this may be an important consideration.[58]

A related but slightly different issue may exist in the BB&W study with respect to import supply. It appears that the analysis is based on the actual real-time supply of imported power and a supply curve slope is estimated for each hour based on day-ahead adjustment bids.[59] The estimated slopes of the import supply curves are not reported but are characterized as extremely inelastic.[60] This is an important and surprising finding. While it might not be too surprising to find that the supply curve for imports into California is relatively inelastic during hours in which there was a

[58] See CalPX Compliance Unit, Price Movements in California Electricity Markets, August 31, 2000 p. 11 (hereafter CalPX I), and CalPXII pp. 55-57. It should be kept in mind that because of the structure of the California market, one might actually observe much of the hydro and thermal generation offered into the PX market at the expected market-clearing price. It would then be difficult to sort out which units were marginal in a cost sense and which units are simply bidding to ensure that they are paid the market-clearing price.

[59] BB&W pp. 24-28, 30, 40.

[60] BB&W p. 30.

general capacity shortage in the WSCC, this is a surprising finding over the long period analyzed by BB&W. It is important to better understand the methodology leading to this conclusion and if the methodology is validated, their finding suggests that there ought to be an inquiry into whether there are features of the California or WSCC markets that give rise to such an inelastic day-ahead supply curve for imports. These aspects of the BB&W analysis are particularly important given the apparently central role of imports in setting PX prices during the summer of 2000.[61] Moreover, the CalPX indicates that the elasticity of import supply in the PX is relatively high, which is inconsistent with the BB&W findings regarding import elasticity in the earlier period.[62]

If we understand the methodology correctly, the slope of the supply curve for import schedules estimated by BB&W based on the day-ahead market is then applied to the real-time import supply quantity. It is noteworthy that in hours in which the real-time price is much higher or lower than the day-ahead PX price, the location of the simulated supply curve for imports may be rather different from either the day-ahead or real-time import supply curve. Indeed, this is one of the issues that ought to be investigated in understanding the apparent inelasticity of import supply in the day-ahead markets, whether a similar inelasticity is seen in the real-time supply curve for imports.[63]

E. Transmission Congestion

BB&W do not attempt to model the impact of transmission congestion on the cost of meeting California load in a competitive market. Instead, they compare the market-clearing price that they estimate, ignoring the effects of congestion, with the hypothetical unconstrained PX price, which is also calculated without regard to the impact of transmission congestion.[64] While this approach greatly simplifies the analysis, it has several relatively subtle implications for the BB&W analysis. First, it must be kept in mind that the fixed schedules for hydro and geothermal generation and the actual import schedules used by BB&W to estimate the competitive price were determined by the actual zonal PX prices, not the hypothetical unconstrained PX price. Thus, if the actual zonal price in a particular period was higher than the hypothetical unconstrained PX price (such as for NP 15 in September 1999), then the actual supply of hydro and geothermal resources within NP15 (and possibly imports from the Pacific Northwest) would be higher than the hydro and geothermal supplies included in the PX supply curve at the unconstrained PX price. This would tend to cause BB&W to estimate a lower "competitive unconstrained PX price" than the actual "unconstrained PX price" even if every thermal generator submitted bids to the PX that were

---

[61] CalPX II pp. 55-57.

[62] See BB&W p. 30 vs CalPX II p. 59-61.

[63] If the day-ahead import supply curve is dominated by must-run resources, or bilateral schedules submitted by scheduling coordinators other than the PX that are price inelastic because of the lack of internal California resources while more price elastic import supply is bid into the hour-ahead market, the slope of the import supply curve estimated from day-ahead adjustment bids may be much steeper than the import supply curve that clears the market in real-time.

[64] BB&W p. 18.

identical to the incremental production costs estimates used by BB&W to define the supply curve. Conversely, if the actual PX price were lower in a zone than the unconstrained PX price used by BB&W, then the supply curve used by BB&W would include less hydro and geothermal resources than the supply curve used to calculate the unconstrained PX price. If hydro and geothermal resources are located largely within NP15, then this consideration would tend to bias the estimate of competitive prices down in periods in which the NP15 price materially exceeded the unconstrained price. This was the case in July, August and September 1999 as the NP 15 price exceeded the unconstrained price by an average of $2/MWh or more, and BB&W simulated a lower hypothetical unconstrained price than the actual hypothetical unconstrained PX price in all of these months. The bias exists on an hour by hour basis, however, and its impact depends on the prices at which the hydro and geothermal resources were bid into the market, so it is difficult to draw reliable conclusions regarding the existence or magnitude of the bias from monthly average data.

A second effect of congestion is that if the transmission congestion were expected day-ahead and expected to exist in real-time, then inter-temporal arbitrage and the segmented structure of California energy markets would have caused thermal generators located in the constrained up regions that entirely lack market power to submit PX bids that exceeded the unconstrained PX price, but instead approximated the actual zonal PX price. Thus, in a world with perfect competition and perfect foresight, the unconstrained PX price would exceed the price simulated by BB&W in any period in which the NP15 or SP15 price exceeded the unconstrained PX price. As with the first effect, it is difficult to assess the practical magnitude of this bias from average monthly data. Unlike the first bias, this bias affects the supply of thermal generation and would be found in both SP15 and NP 15.

A third effect of congestion relates to intra-zonal congestion. Until the implementation of amendment 26 in early 2000, much intra-zonal congestion was managed through the call of RMR contracts, after the close of the PX market. In this non-locational bidding structure, firms lacking market power but having perfect foresight would be expected to bid their capacity into the market at the locational market-clearing price. Firms with potential market power but subject to RMR restrictions would, again with perfect foresight, offer their capacity into the market in periods with transmission congestion at the lower of the locational market-clearing price or the mitigated price under their RMR contracts, and would be called by the ISO out of merit at the RMR contract price to meet real-time load. This is recognized by BB&W who note the potential for RMR contract calls to cause the actual PX price to be lower than the simulated PX price curve used in their analysis.[65]

Although neither the BB&W nor MSC papers discuss the implications of prescheduling of RMR generation subsequent to the implementation of Amendment 26, this system would also in effect cause the unconstrained PX supply curve to be lower and to the right of the supply curve estimated by BB&W and the MSC. Like the other biases, the magnitude of the impact of RMR generation on price is variable and hard to quantify without reference to data on actual RMR calls.

[65] BB&W p. 29

F. One-Part Bids

We discussed above the complexity of accounting for start-up and minimum load costs in understanding the bidding of competitive generators under a one-part bidding system (see section II D). The BB&W analysis excludes start-up and minimum loads costs in defining the simulated competitive supply curve and in simulating the estimated market-clearing price.[66] While acknowledging that these costs are not sunk and must enter into commitment decisions, they observe that "it is not at all obvious for a price-taking profit-maximizing firm to prorate such costs into its supply bids. In fact, it is relatively easy to construct examples where it would clearly not be optimal to do so."[67] There are two problems with this rationale for excluding start-up and no-load costs. First, even if the appropriate strategy were not to prorate costs, this would not justify simply ignoring the costs. Second, while BB&W apparently believe that their example indicates that the optimal bid would always be less than the prorated cost, this is not the case. As outlined in section II D above, the optimal bid in these circumstances for a generator that lacked market power and expected to be able to recover its start-up and no-load costs at expected day-ahead market prices would be less than the prorated cost, and likely equal to marginal production cost. On the other hand, the optimal bid of a generator that lacked market power but did not expect to be able to recover its start-up and no-load costs at day-ahead prices would likely be higher than its incremental production costs plus prorated start-up costs. Indeed, the prorated cost assumption might be much too conservative.

BB&W also argue that the actual start-up costs were small, probably less than $30 million/year.[68] The basis for this estimate is not provided, but even if correct, this estimate does not necessarily imply a small impact of start-up costs on the competitive price level. First, if the proportion of start-up costs that a potentially marginal generator did not expect to recover in real-time energy prices was 1/3, then such generators would in aggregate have included $10 million of start up costs in their day-ahead supply offers. If the average unit size of these marginal units were 250MW then there would be $40,000/MW year of start-up costs to recover in energy bids over the year or about $4.50/MWh, which is large relative to the price differences estimated in these studies. Of course, not all of the units incurring start-up costs would necessarily be marginal in the hours in which their bids were accepted, and the relevant percentage of costs not expected to be recoverable in real-time prices might be lower than 1/3, both of which would reduce the dollar impact of ignoring these costs. Nevertheless, the point is that it cannot simply be assumed that the impact of start-up costs on the one-part bids of competitive suppliers in the California market is immaterial.

Second, the $30 million dollar start-up cost figure apparently pertains to the actual level of start-up costs, not the level that would have been incurred had generators bid into the PX as if they did

[66] BB&W p. 22.

[67] BB&W p. 22.

[68] BB&W p. 29.

not have start-up costs. The implicit level of start-up costs in the schedules underlying the BB&W simulated supply curve is not known, but the omission of start-up costs from the simulation would tend to cause low variable cost, high start-up cost, units to be used to meet peak load, rather than relying on higher variable cost peaking units.

The reality is that simple models that do not account for start-up and no-load costs cannot accurately simulate the competitive level of prices under a one part bid, hour by hour market such as that operated by the California PX. A better alternative would be to assume that market participants were able to replicate the results of centralized unit commitment through their individual one part bids and self-commitment and thus to estimate market prices based on the actual multi-part cost functions of the generators. [69] As discussed above, this analysis of market prices would still be biased towards finding market power because it would tend to attribute to market power the inefficiency of the California market, but the estimates would be less biased than those provided by a methodology which simply assumes that start-up and no-load costs do not exist.[70]

G. Data Issues

There are also a variety of data problems with the BB&W study that appear likely to generally lead to an underestimate of the competitive price level.

1. Outages

BB&W note that the generation supply curve they estimate is not adjusted for the impact of scheduled outages. [71] BB&W reason that this is appropriate because the scheduling, and duration of maintenance and other outages is a strategic decision and thus could be used to withhold capacity from the market. Moreover, they observe that they "would expect scheduled maintenance to take place in the autumn, winter and spring months, which is the time period over which we find little, if any, market power."[72] The MSC study, however, finds market power in September-December 1999, and January and May 2000. [73] Whether scheduled outages affect prices in the shoulder months is a matter of chance. While the demand for capacity is generally lowest in the shoulder months, shortages occasionally occur because of unusual weather conditions or forced outages. Simply assuming that no maintenance outages are ever necessary

[69] Simulation models like GE MAPS could do this, as well as taking account of transmission constraints.

[70] It is noteworthy that the authors did not take such an approach since the incremental heat rate data were derived from the CEC's GE Maps database. GE-MAPS could be used to simulate unit commitment and prices taking account of start-up and no-load costs. This approach would also have enabled simulation of the effect of locational constraints on RMR calls and modeling of reserve requirements.

[71] BB&W p. 21.

[72] BB&W p. 21.

[73] MSC p. 17.

removes this source of capacity shortages from the supply curve and thus on average underestimates the competitive price in the shoulder months.

While the direction of impact is uncertain, it should be noted that both the month–to-month simulation results and the longer-term pattern of simulated market prices depends on whether the assumptions regarding forced outage rates converge to the actual rate observed for California thermal generators. The NERC forced outage data used in the BB&W study is a reasonable starting point, but it is not necessarily applicable to the California units or under the utilization conditions prevailing in the summer of 2000. Although outages could be a method of exercising market power, this is not necessarily the case. As an alternative to the use of simulated outage data in the MSC analysis, that analysis could be based on actual outage data, which should be available.

2. Emission allowance costs and restrictions

It does not appear that the BB&W analysis took any account of emission allowance costs or operating constraints. BB&W suggest that this is not necessary because allowance costs were less than $400/ton.[74] This may have been the case during the 1998-1999 period but recent data, suggest NOx emission costs of around $20,000 a ton early in the summer of 2000, rising to around $70,000/ton later in the summer of 2000.[75] It would therefore appear that MSC conclusions regarding the exercise of market power during the summer of 2000 need to be tempered with the recognition that the analysis omits an important incremental cost element. Moreover, to the extent that individual generators are subject to absolute emission limits, these limits would be ignored in the BB&W and MSC analyses. These limits would likely be particularly important during sustained levels of high demand such as that which prevailed in the summers of 1998 and 2000.

While emission allowance costs would likely not dramatically affect the incremental economics of new low-emission units, it is important to keep in mind that at high load levels, the marginal unit will not be a new low-emission unit. For older units, with emission rates in the range of 1-4 lbs/MWh that may have been on the margin during peak periods during the summer of 2000, the emission allowance costs reported by the PX [76] imply incremental emission allowance costs perhaps in the range of $10-$40/MWh in June and $35 to $140/MWh in July and August. The California PX has suggested some units might have NOx emission rates as high as 10 lbs/MWh.[77] Combined with high gas and oil prices, the incremental production costs of many fossil peaking units would have been extremely high during the summer of 2000 and especially in August 2000.

---

[74] BB&W p. 21.

[75] Cal PX I p. 9, CalPX II pp. 28-30. Cantor Fitzgerald kindly allowed us access to confidential data on their website (www.cantor.com) to allow us to confirm the data cited by the PX. The transactions are not dated, so it is difficult at this point in time to confirm precisely when allowance prices rose.

[76] CalPX I p. 9, CalPX II p. 29.

[77] CalPX II p. 30.

In a centrally coordinated market for energy and reserves, such extremely high operating cost units would generally be scheduled to provide reserves and to only operate during emergency conditions. It is not known to what extent such high-cost peaking units were actually operating and setting real-time prices in California electricity markets during the summer of 2000,[78] either as a result of the general inefficiency of the California market or as a result of energy and capacity shortages.

3. Fuel price data

BB&W base their estimates of market prices on weekly average natural gas spot prices at the PG&E city gate and California-Arizona border, adjusted in some manner for the gas distribution rates of the gas utility serving each generator.[79] Although the prices used in the BB&W analysis are averaged over the week, minor day to day variations in gas prices, unlike capacity shortages, should average out without materially affecting the results.[80]

The BB&W paper states that the competitive analysis is based on a #2 fuel oil price of $2.98/Million btu.[81] Based on a 5.825 Million-btu/bbl rate, this translates into a price of $17.36/bbl. This cost figure originates in a single oil delivery in February 1998. The source BB&W rely on for these oil cost data actually contains very little information because FERC Form 423 data is not reported for the small peaking units that would use these fuels in California. Monthly average spot prices for LA Harbor for the period June 1998 to August 2000 are portrayed in Table 14. It can be seen that the #2 fuel oil price used to simulate the competitive level of electricity prices is roughly on track during mid 1998, high during late 1998 and early 1999 and then falls increasingly below market levels and is far below market prices during the summer of 2000.

Similarly, the BB&W paper states that the competitive price analysis is based on a jet fuel price of $3.29/Million btu based on a MAPS data set.[82] Based on a 5.670 Million-btu/bbl, this translates

---

[78] It should be kept in mind that even if these high cost units were never scheduled in day-ahead markets, their costs would affect day-ahead prices if the high cost units operated in real-time, determining both actual and expected real-time prices.

[79] BB&W p. 39. The accuracy of the incremental production costs for the gas fired generation also depends on the accuracy of the distribution charges included in those costs which was not described in the BB&W papers.

[80] There is also some subtlety in the assumptions regarding gas prices. Real-time prices would be determined on the margin not by day-ahead gas prices but by swing gas prices. These gas prices could be higher or lower than day-ahead gas prices. If these deviations were symmetric, their impact on real-time vs day-ahead electricity prices would be symmetric.

[81] BB&W p. 39. The reference dates from earlier versions of this paper and pertains to data published in 1998, which as noted above actually pertain to costs in February 1998. It is not clear whether the current footnote accurately reflects the data used in the study.

into a price of $18.65/bbl. Monthly average spot prices for LA harbor the period June 1998 to August 2000 are also portrayed in Table 14. It can be seen that the jet fuel price used to simulate the competitive level of electricity prices also falls increasingly below market levels after June 1999 and is far below market levels during the summer of 2000.

These are material differences, as a difference of $15.00/bbl in June 2000 would translate into a difference of $30/MWh for a unit with a 12,000Btu/kWh heat rate and $38/MWh for unit with a 15,000 Btu/kWh heat rate. As with emission costs, the impact of these data limitations on the BB&W results is hard to evaluate because it is difficult to assess how often these units were on the margin in the hypothetical study or would be expected to be on the margin in the real-world.

4. Negative Market Power

The BB&W and MSC papers simulate prices that are higher than the actual unconstrained PX prices in many hours and average lower than the actual unconstrained PX price for several months. The cause of these instances of "negative market power" has implications for how BB&W measure the exercise of market power and for assessing biases in the model. [83]

BB&W note two considerations that in their view would cause the estimated price to sometimes be higher and sometimes lower than the cost curve bid into the PX. These are mistakes in bidding minimum load blocks into the PX in hours when the unit is expected to be economic but turns out to be uneconomic, and errors in the cost data for generating units (apparently referring to errors in heat rate data).[84] At least three other sources of random error probably exist in the BB&W analysis. First, there are differences between actual and simulated fuel costs. The BB&W cost analysis is based on weekly average gas prices rather than daily spot gas prices. This will cause their supply curve to be too low on some days and too high on others. Moreover, we have noted above that their petroleum fuel costs at times differ widely from market prices. Second, the use of real-time load to estimate day-ahead prices will cause day-to-day and month-to-month variations as discussed in Section III C above. Third, variations between actual and simulated outage rates will cause day-to-day and month-to-month variations between the actual and simulated prices. These considerations are likely to cause a significant degree of variation around the expected price and it is important, as BB&W and the MSC have done, to net hours in which "positive" and "negative" market power are measured. The observation that the BB&W and MSC studies do not identify material market power in some months does not establish that the study is unbiased as these sources of bias may have merely been offset by these month to month variations in these other sources of error.

[83] BB&W pp. 28-29, 34.

[84] BB&W pp. 28-29.

Table 14

	LA Spot Jet Fuel	LA Spot No. 2 Diesel
	$/bbl	$/bbl
6/1/98	16.85	17.07
7/1/98	18.40	18.04
8/1/98	19.48	17.78
9/1/98	19.46	18.43
10/1/98	19.54	18.40
11/1/98	17.73	17.67
12/1/98	16.05	15.61
1/1/99	17.12	15.99
2/1/99	15.70	15.21
3/1/99	19.07	20.22
4/1/99	23.37	21.21
5/1/99	20.65	20.07
6/1/99	26.47	23.67
7/1/99	25.58	25.87
8/1/99	27.21	27.29
9/1/99	28.21	25.83
10/1/99	28.00	26.68
11/1/99	30.44	30.25
12/1/99	32.16	30.01
1/1/00	37.13	32.58
2/1/00	36.01	33.74
3/1/00	39.82	37.84
4/1/00	32.72	31.09
5/1/00	33.68	31.65
6/1/00	34.27	34.43
7/1/00	35.73	35.90
8/1/00	41.89	39.92
9/1/00	47.89	47.01
10/1/00	45.71	46.50

Note: the Monthly prices are calculated as the simple average of the daily closing spot prices.
Source: National Energy Information Center of the Energy Information Administration.

IV. Conclusions

Given the importance of the California electricity market itself, and the lessons that might apply to other electricity markets, the events of the summer of 2000 are troubling and deserve extensive analysis. Unfortunately, the design problems already acknowledged in the California market greatly complicate the process of untangling what has been happening. Since policy recommendations depend critically on what is wrong, it is essential to pin down the answer better than it has been so far. The work of BB&W and the MSC stands virtually alone in attempting to provide an analytical and empirical foundation for a judgment that the high prices in the California electricity market during the summer of 2000 are accounted for by a significant exercise of market power. Examination of the evidence they provide suggests that these studies are really the beginning of the inquiry, not the end of the story.

As suggested by the discussion above, other parts of the story that will need thorough examination in future studies are the role of: i) capacity shortages in California and elsewhere in the WSCC; ii) market separation, pay-as-bid pricing and related inefficiencies in the California market structure; iii) the supply elasticity of imports into California; and iv) the rising cost of emission allowances within California. In addition, it will be necessary for these future studies to carefully take account of start-up and no-load costs, the impact of inter- and intra-zonal congestion, and non-thermal supply offers.